# Predistortion of GaN Power Amplifier Transient Responses in Time-Division Duplex Using Machine Learning

Arne Fischer-Bühner, *Graduate Student Member, IEEE*, Lauri Anttila, *Member, IEEE*, Alberto Brihuega, Manil Dev Gomony, *Member, IEEE*, and Mikko Valkama, *Fellow, IEEE*

***Abstract*—The extensive use of time-division duplexing in 5G and 6G poses a challenge to the linear operation of the power amplifiers (PAs) in radio base stations. Particularly with gallium nitride (GaN) technology, the PAs can produce strong transient behavior when resuming from an idle state, which degrades the first few transmitted symbols. This article proposes a novel machine learning technique to model and compensate the PA gain transient, based on a lightweight, low-rate recurrent model. Our RF measurements at 3.6 GHz examine the joint application of transient compensation and predistortion of short-term effects and show a successful mitigation of both types of distortion.**

***Index Terms*—Digital predistortion, gallium nitride (GaN) power amplifier (PA), long-term memory effect, time-division duplex (TDD), transient response.**

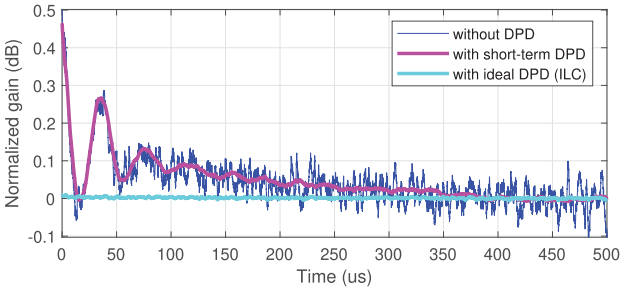


Fig. 1. Gain response of a GaN PA when resuming operation in TDD, without, with short-term DPD [7], and with ideal DPD.

## I. Introduction

THE advent of 5G new radio (NR) has brought about significant changes to the wireless radio access network [1]. Particularly, new frequency bands and a flexible numerology and resource allocation were introduced to increase spectrum utilization and support a broad range of use cases. Among these changes, time-division duplexing (TDD) has emerged as a dominant duplexing technique for the unpaired frequency bands above 3 GHz, with a flexible assignment of time resources for uplink (UL) and downlink (DL) [2]. While the flexibility is key to achieving the various 5G and future 6G targets, the rapid alternation of UL/DL poses a significant challenge to the design of the radio base stations. For the analog front-end, especially the radio frequency (RF) power amplifiers (PAs), interruption of the transmission results in a transient gain response when resuming operation, which may introduce considerable distortion to the first transmitted symbols. Capable mitigation techniques are thus needed to avoid compromising the envisioned benefits of dynamic TDD.

To achieve high-quality transmission in radio base stations, and to comply with regulatory requirements, digital predistortion (DPD) is typically applied to mitigate the distortions occurring in the RF PA [3]. DPD linearization of PA nonlinearity and *short-term* memory behavior in a static transmission setting has been extensively studied. Reported methods range from efficient look-up table schemes [4], to more accurate Volterra-based models [5], and neural network (NN)-based techniques [6], [7], [8]. These DPD models typically operate in a per-sample mode to address the memory effects in a 10–50-ns time window. In contrast, a PA transient response has a *much longer time span*, from $\mu$s to 100 ms. Consequently, transient effects cannot be captured by the short-term memory of typical DPD models [9]. Prior studies used single-pole IIR filters or moving average power estimates to model gain transients, which were either manually tuned [10], [11] or extracted based on physics-based consideration and specific laboratory measurements [12], [13]. However, existing approaches limit themselves to only first-order responses and lack the possibility to automate filter identification.

In this article, we propose a data-driven machine learning (ML) technique to compensate both decaying and oscillating PA transient behavior. A recurrent model is used at a reduced sampling rate to learn the large-scale gain variations. We then apply the gain model together with a short-term NN DPD to compensate for different types of distortion concurrently for optimized overall performance. This article is organized as follows. Sections II and III develop a gain model with ML-based tuning. Section IV studies different possibilities to incorporate the gain model with the short-term DPD.

Received 28 February 2025; accepted 6 April 2025. Date of publication 30 April 2025; date of current version 9 June 2025. This work was supported in part by EU Horizon 2020 under the Marie Skłodowska-Curie under Grant 860921; and in part by the Academy of Finland under Grant 332361, Grant 345654, Grant 352754, and Grant 357730. *(Corresponding author: Arne Fischer-Bühner.)*

Arne Fischer-Bühner is with Nokia Bell Labs, 2018 Antwerp, Belgium, and also with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland (e-mail: arne.fischer@nokia-bell-labs.com).

Lauri Anttila and Mikko Valkama are with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland.

Alberto Brihuega is with Nokia Mobile Networks, 90620 Oulu, Finland.

Manil Dev Gomony is with Nokia Bell Labs, 2018 Antwerp, Belgium, and also with the Department of Electrical Engineering, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands.

This article was presented at the IEEE MTT-S International Microwave Symposium (IMS 2025), San Francisco, CA, USA, June 15–20, 2025.

Digital Object Identifier 10.1109/LMWT.2025.3561227

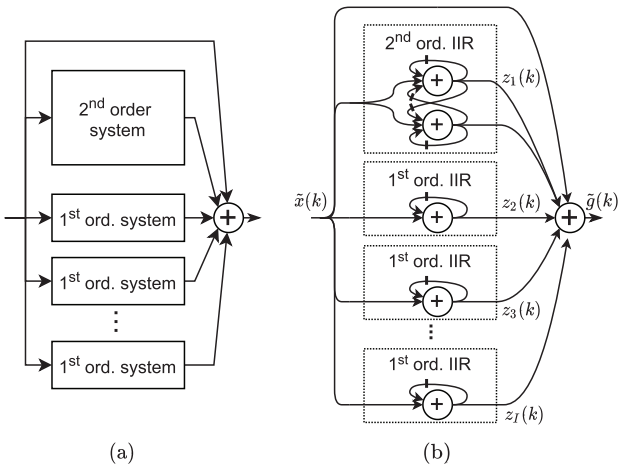


Fig. 2. (a) Parallel IIR system structure and (b) recurrent network model graph, realizing the IIR system of (a). Black bars denote sample delays.

RF experimental results at 3.6-GHz band are presented in Section V, and Section VI concludes this article.

## II. Modeling of the Gain Transient Response

Modern 5G radio base stations primarily rely on gallium nitride (GaN) power amplifiers, due to their suitability for high power operation at higher frequencies [14]. GaN PAs typically show a strong transient behavior in response to changes in the operation power, sometimes referred to as long-term memory effects [9]. Fig. 1 shows the measured gain response of a GaN PA, operating in the C-band (3.6-GHz carrier center frequency), after resuming the operation from idle state. A short-term NN DPD, based on [7], successfully corrects for the largest part of the PA gain fluctuations; however, the long-term gain transient response remains unaffected by the short-term DPD, whereas an ideal DPD would produce a flat gain.

Several mechanisms are identified as potential sources for the transient. First, thermal effects can affect the gain, as PAs typically have higher gain in a colder state. During start-up, a gain drop can be observed as the PA heats up [12], [15]. Second, trapping of electrons in the GaN substrate can reduce the gain of the power amplifier as electrons are trapped in the uncharged substrate [9], [13]. Third, an unstable supply and biasing can impact the gain or nonlinearity of the PA in pulsed operation [16]. Depending on the specific PA unit, one mechanism may dominate the overall response or the various effects may coexist. Yet, each of the individual thermal and electrostatic mechanisms may be fundamentally described as a first-order (decaying) or second-order (damped oscillating) system.

To develop this into a machine learning model, we assume that the PA gain behaves sufficiently linear with respect to the trap charge, temperature, and supply and bias voltage. We then interpret the gain transient as a step response that can be explained by a superposition of first- and second-order systems. Consequently, we can model each subsystem individually using infinite impulse response (IIR) filters. This approach is depicted in Fig. 2(a). When identifying the model, we thus seek a suitable set of digital first- and second-order IIR filters, which resemble the overall transient behavior when combined. Fig. 3 illustrates the approach, where several IIR filters were manually tuned and combined to match the transient response.

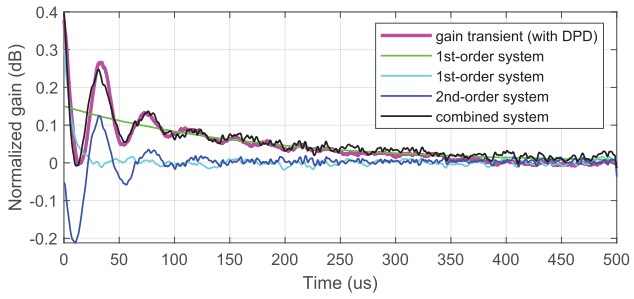


Fig. 3. Combination of different manually tuned first- and second-order systems to resemble the measured gain transient.

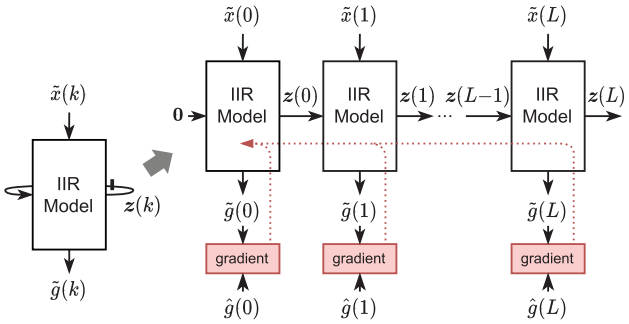


Fig. 4. Unfolding of the recurrent structure into a sequence-to-sequence model for training. The red arrows indicate the flow of the gradients during BPTT.

## III. Machine Learning Gain Transient Model

Given the limitations of manual filter tuning, we propose a data-driven method for automated tuning. We formalize our assumption of parallel systems as a network graph, shown in Fig. 2(b), with the first- and second-order IIR filters modeled as recurrent structures akin to recurrent neural network (RNN) without nonlinear activation or gating mechanisms. The IIR filters are given with

$$\boldsymbol{z}_i(k) = \boldsymbol{a}_i \tilde{x}(k) + \boldsymbol{b}_i \boldsymbol{z}_i(k-1) \tag{1}$$

where $\boldsymbol{a}_i \in \mathbb{R}^{N_i \times 1}$ and $\boldsymbol{b}_i \in \mathbb{R}^{N_i \times N_i}$ are the coefficients, and $\boldsymbol{z}_i \in \mathbb{R}^{N_i x 1}$ is the output of the $i$th IIR filter with order $N_i$. The outputs of the filters are combined linearly to model the PA gain $\tilde{g}(k)$

$$\tilde{g}(k) = c_0 \tilde{x}(k) + \sum_{i=0}^{I} \boldsymbol{c}_i z_i(k) \tag{2}$$

where $\boldsymbol{c}_i \in \mathbb{R}^{1 \times N_i}$, $N_0 = 1$.

To tune the coefficients, we apply data-driven machine learning, i.e., coefficients are initialized, followed by iterative gradient adaptation with respect to a baseline training dataset $\hat{g}(k)$. We derive gradients with a back-propagation-through-time (BPTT), which is typically used to optimize RNNs [17]. Therein, as depicted in Fig. 4, the recurrent structure is unfolded for $L$ time steps, to construct a sequence-to-sequence model. In this mode, the model processes a full sequence at a time, and gradients for optimizing the coefficients can be estimated across the unfolded recurrent connections of the model. The coefficients, however, remain shared for the unfolded instances.

Two challenges for the optimization need to be addressed.

1) The sampling rate of the baseband signal is very high in relation to long-term transient effects. At this rate, the

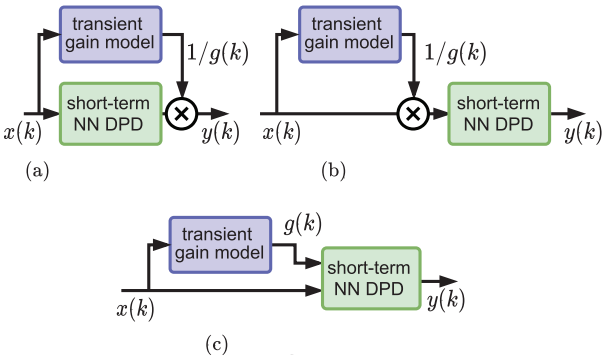


Fig. 5. Different configurations for joint short-term and transient DPD: (a) gain correction is applied in front of the short-term DPD, (b) gain correction is applied at the output of the short-term DPD, and (c) short-term DPD is augmented with the transient gain prediction as an additional. (a) Configuration A. (b) Configuration B. (c) Configuration C.

respective poles of the digital IIR filters are very close to one and the filter responses are sensitive to even small changes in the coefficients. Moreover, for BPTT, the structure would require extensive unfolding, rendering the training complexity impractical. To solve this issue, we propose to train and operate the model at a significantly reduced rate. Thus, poles and respective filter coefficients become more manageable, and the model can easily be unfolded to perform BPTT with the full gain response. This is expressed through $\tilde{x}(k)$ and $\tilde{g}(k)$, which are the low-rate signals corresponding to $x(k)$ and $g(k)$. For downconversion, we use a windowed average of the gain. Upconversion is achieved by replicating the filter output between the low-rate samples.

2) Gradient-based tuning of the second-order IIR filter is not robust for a damped oscillating response. With a random coefficient initialization, the optimization may be slow, or converge to a suboptimal minimum or plateau. To overcome this, we propose to initialize the coefficients of the second-order IIR structure with a specific initial configuration given by

$$\boldsymbol{a}_i^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b}_i^{(0)} = \begin{pmatrix} 0.9 & 0.1 \\ -0.1 & 0.9 \end{pmatrix} + \mathcal{N}_0 \tag{3}$$

where $\mathcal{N}_0$ is a small noise component to randomize initialization. The asymmetric initialization of the recurrent paths enables the IIR to reliably fit an oscillating response. Zero initialization of $\boldsymbol{a}_i$ resolves complications with exploding gradients in the initial training phase, which is common in RNNs without gating. Accordingly, we also initialize first-order IIR structures with

$$\boldsymbol{a}_i^{(0)} = 0 \text{ and } \boldsymbol{b}_i^{(0)} = 0.9 + \mathcal{N}_0. \tag{4}$$

## IV. Joint Short-Term and Transient DPD

To actually compensate PA gain transients, we use the transient gain model, developed in Section III, for predistortion. Straightforwardly, the transmit signal can be multiplied with the inverse of the modeled gain, $1/g(k)$. However, since short-term DPD should be applied together with the transient compensation, the question is raised on how to properly combine the two DPDs. We examine three different possibilities depicted in Fig. 5(a)–(c) to achieve transient correction in conjunction with an NN-based DPD.

In configuration A, we apply the transient correction by multiplying the output of the short-term DPD with the inverse of the modeled gain, as

$$y_{\mathrm{A}}(k) = 1/g(k)\, f_{\mathrm{DPD}}(x(k)). \tag{5}$$

Variant B applies the gain correction to the input signal of the short-term DPD. This is expressed as

$$y_{\mathrm{B}}(k) = f_{\mathrm{DPD}}(x(k)/g(k)) \tag{6}$$

where $f_{\mathrm{DPD}}$ denotes the operation of the short-term NN DPD. Configuration C augments a short-term DPD by providing the transient gain estimate as additional input to the DPD. We express this as

$$y_{\mathrm{C}}(k) = f_{\mathrm{DPD}}(x(k), g(k)). \tag{7}$$

This enables the NN DPD to perform joint transient correction and short-term memory DPD. The performance of all the three configurations is compared and discussed in the next following results section.

## V. RF Measurement Results

To realize the short-term DPD, we use the phase-normalized time-delay NN from [7] with an input memory depth of 12 and 1517 model coefficients. For configuration C, the additional single NN input is provided in parallel with the phase normalized time delay inputs. To train the short-term NN, we use the iterative learning control (ILC) [18] derived ideal DPD signal as a training baseline. In all the three configurations, the short-term DPD is trained with a perfect gain correction in place, i.e., we apply the measured transient during short-term NN training, according to the selected configuration. The gain model is trained separately using the measured transient. To validate the proposed methods, we used a 5G compliant test signal with an instantaneous bandwidth of 100 MHz in the C-band (3.6 GHz). In the TDD test signal, alternating slots are allocated for UL/DL. A subcarrier spacing of 60 kHz was chosen, and the resulting slot duration is 0.25 ms. The PA in our measurements is a line-up of a driver (*WSGPA01*) and an asymmetric GaN Doherty PA module (*WS1A3640*) on an evaluation fixture with static biasing.

A transient gain model with one first-order IIR filter and one second-order IIR filter was found sufficient to model the gain response of the test case. The transient model uses a sampling rate that is 512 times lower than the DPD sampling rate of 491.52 MHz. The gain model and the short-term NN DPD were trained on a single DL slot. We then tested the linearization capability on a separate waveform with four slots, two for UL and two for DL, respectively. The waveform repeatedly transmitted, and the results are a snapshot of a single repetition. Fig. 6 shows the gain response during a single DL slot, a first-order IIR modeling attempt, the modeled gain response using the proposed ML technique, and the resulting linearized gain when applying both the transient and short-term DPD. Since the transient is dominated by an oscillating effect, the first-order IIR fit can only follow the average of the transient, but does not realize a good fit of the transient. The proposed ML optimized IIR instead closely captures the oscillating effect and enables linearizing the gain response.

Fig. 7 shows the measured per-symbol error vector magnitude (EVM) for the three compared DPD configurations.

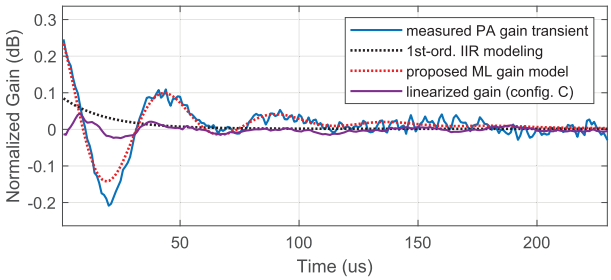


Fig. 6. Measured PA gain response, the modeled gain transient, and the linearized gain in a TDD scenario with an alternating UL/DL slot allocation.

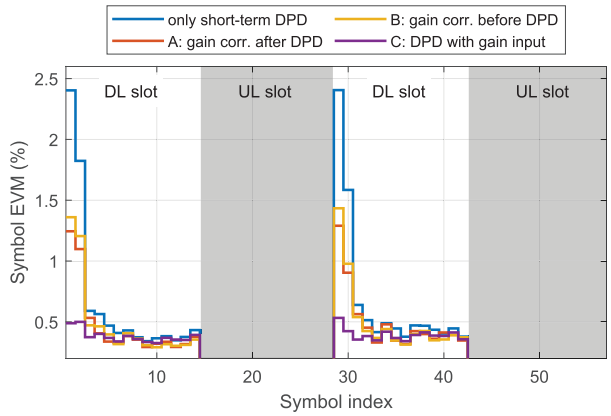


Fig. 7. Symbol EVM at the PA output for an alternating UL/DL configuration.

Without transient DPD, the EVM of the first two symbols of each DL slot is notably degraded due to the transient gain response, while the remaining symbols are properly linearized by the short-term DPD alone and thus have a low EVM. All three configurations achieve a better linearization of the gain transient, compared to the short term DPD baseline, and improve the EVM of the first two OFDM symbols to some extent. However, configurations A and B still only partially mitigate the transient and the EVM of the first two symbols remains significantly increased. Configuration A performs slightly better compared with configuration B. Finally, configuration C achieves a consistently low EVM for all the symbols. The gain transient is successfully compensated. As the short-term NN DPD is aware of the gain variation, it can incorporate any impact of the gain transient on the short-term effects.

## VI. Conclusion

This article proposed a technique for compensation of PA gain transient effects. A recurrent network model was developed to model the transient gain characteristic, which can be optimized using ML. We studied different possibilities to incorporate the ML optimized gain model with a short-term DPD, and we found that augmenting a short-term DPD with the prediction of the transient-gain model allows to mitigate the transient gain response in conjunction with effective short-term DPD.

## References

[1] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*. New York, NY, USA: Academic, 2020.

[2] K. Chuang, H. Yektaii, N. Outaleb, A. Raslan, S. Bhal, and P. Forbes, "Towards sustainable networks: Attacking energy consumption in wireless infrastructure with novel technologies," *IEEE Microw. Mag.*, vol. 24, no. 12, pp. 44–59, Dec. 2023.

[3] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1880–1890, May 2017.

[4] P. L. Gilabert, A. Cesari, G. Montoro, E. Bertran, and J.-M. Dilhac, "Multi-lookup table FPGA implementation of an adaptive digital predistorter for linearizing RF power amplifiers with memory effects," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 2, pp. 372–384, Feb. 2008.

[5] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[6] 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.

[7] A. Fischer-Bühner, L. Anttila, M. D. Gomony, and M. Valkama, "Phase-normalized neural network for linearization of RF power amplifiers," *IEEE Microw. Wireless Technol. Lett.*, vol. 33, no. 9, pp. 1357–1360, Sep. 2023.

[8] T. Kobal and A. Zhu, "Digital predistortion of RF power amplifiers with decomposed vector rotation-based recurrent neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4900–4909, Nov. 2022.

[9] J. L. Gomes et al., "The impact of long-term memory effects on the linearizability of GaN HEMT-based power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 2, pp. 1377–1390, Feb. 2022.

[10] J. P. Aikio, J.-J. Toivanen, T. Kolmonen, A. Brihuega, A. Pärssinen, and T. Rahkonen, "Effects and modeling of GaN PA distortion in TDD mode using 5G NR signal," in *Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits (INMMIC)*, Nov. 2023, pp. 1–3.

[11] A. S. Tehrani, T. Eriksson, and C. Fager, "Modeling of long term memory effects in RF power amplifiers with dynamic parameters," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2012, pp. 1–3.

[12] S. Boumaiza and F. M. Ghannouchi, "Thermal memory effects modeling and compensation in RF power amplifiers and predistortion linearizers," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 12, pp. 2427–2433, Dec. 2003.

[13] F. M. Barradas, L. C. Nunes, T. R. Cunha, P. M. Lavrador, P. M. Cabral, and J. C. Pedro, "Compensation of long-term memory effects on GaN HEMT-based power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3379–3388, Sep. 2017.

[14] D. Y. C. Lie, J. C. Mayeda, Y. Li, and J. Lopez, "A review of 5G power amplifier design at cm-Wave and mm-Wave frequencies," *Wireless Commun. Mobile Comput.*, vol. 2018, no. 1, pp. 1–16, Jan. 2018.

[15] K. Ohgami, Y. Murao, and T. Kaneko, "Transient thermal response impact of 3 GHz GaN HEMT amplifier on TDD LTE spectrum and its improvement based on a thermal equivalent circuit approach," in *Proc. IEEE Compound Semiconductor Integr. Circuit Symp. (CSICS)*, Oct. 2016, pp. 1–4.

[16] C. G. Tua, T. Pratt, and A. I. Zaghloul, "A study of interpulse instability in gallium nitride power amplifiers in multifunction radars," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 11, pp. 3732–3747, Nov. 2016.

[17] P. J. Werbos, "Generalization of backpropagation with application to a recurrent gas market model," *Neural Netw.*, vol. 1, no. 4, pp. 339–356, Jan. 1988.

[18] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative learning control for RF power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 9, pp. 2778–2789, Sep. 2016.