Digital Predistortion of RF Power Amplifiers With Decomposed Vector Rotation-Based Recurrent Neural Networks

Tugce Kobal, *Graduate Student Member, IEEE*, and Anding Zhu, *Fellow, IEEE*

Abstract—In this article, we present a novel decomposed vector rotation (DVR)-based recurrent neural network behavioral model for digital predistortion (DPD) of radio frequency (RF) power amplifiers (PAs) in wideband scenarios. By representing memory terms of DVR with recurrent states and redesigning the piecewise modeling, we propose a novel recurrent DVR scheme. To ensure stable operation and enhanced modeling accuracy, we integrate the recurrent DVR into the gated learning mechanism of the modified Just Another NETwork (JANET). Experimental results confirm that the proposed DVR-JANET model provides much improved linearization performance with significantly reduced model complexity compared with the recent existing models.

Index Terms—Behavioral modeling, digital predistortion (DPD), neural network, power amplifier (PA), recurrent neural network.

I. INTRODUCTION

WITH the advance of wireless standards, the fifth-generation (5G) communication systems are adopting higher carrier frequencies and wider signal bandwidth. While such an adaptation serves to high capacity communication, modern radio frequency (RF) systems inevitably face significant challenges in maintaining high linearity while minimizing power consumption [1]. Genuinely, power amplifiers (PAs) account for the majority of the power and cause foremost nonlinear distortion in RF systems [2]. Owing to its allowing PAs to be operated at higher drive levels and compensating the nonlinear distortions, digital predistortion (DPD) is widely deployed in wireless base stations [3]. In the literature, numerous DPD models have been proposed, such as memory polynomial (MP) [4], generalized MP (GMP) [5], envelope MP (EMP) [6], dynamic deviation reduction (DDR) [7], [8], and so on. In fact, the existing models have been competent in practice. However, with constantly increasing signal bandwidth, the PA behavior is now associated with more sophisticated nonlinear characteristics and memory effects. While endeavoring to model such complex systems, the existing models may fail to meet the required linearization performance.

Relying on the impactful success of deep learning and its enormous potential for improvement, neural network learning emerges an appealing alternative in PA modeling. Recently, many neural network-based DPD models have been proposed [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]. When deploying a deep learning model, selecting a proper network structure to accurately model nonlinear dynamics of the PA becomes significantly challenging. In the existing works, DPD models are mainly built on traditional feedforward networks learning with memory samples. With longer memory effects, however, such models require a quite large number of memory samples, which intrinsically contributes to the complexity of the model. Recurrent models can involve an infinite number of past samples into the learning and, therefore, potentially can lead to higher accuracy [21], but the conventional recurrent networks suffer from stability issues. Some recent works have proposed recurrent DPD models [16], [17], [18], [19], but none of the revised networks concentrate on the search for the most suitable network structure. In [22], a useful link between the lightweight Just Another NETwork (JANET) [23] and the physical behavior of the PA is established. After adopting a new model structure, the phase-gated recurrent neural network-based DPD model, phase-gated JANET (PG-JANET), proposed in [22], achieved remarkable improvement in modeling accuracy and linearization performance. However, a large number of coefficients are still required when linearizing wideband PAs.

In this article, we significantly improve our prior work in [22] and propose a novel decomposed vector rotation (DVR)-based recurrent network to better model the complex nonlinearity and longer memory effects of the PA and, therefore, to ensure a more practical design. Based on the methodical analysis of recurrent operation and DVR model structure [24], we find that it is possible to represent memory terms within the model with recurrent states. By rearranging piecewise modeling according to recurrent learning, we build a novel recurrent DVR scheme to adapt the flexible and powerful nonlinear modeling ability of the DVR model to recurrent neural network learning. To better map the outcome of novel DVR layer onto overall modeling, we also carefully redesign the recurrent JANET unit. Experimental results confirm that the proposed DVR-based recurrent model not only performs superior to the conventional DPD models but also reduces the model complexity significantly.

The rest of this article is organized as follows. Section II reviews PA modeling with PG-JANET. Section III explains
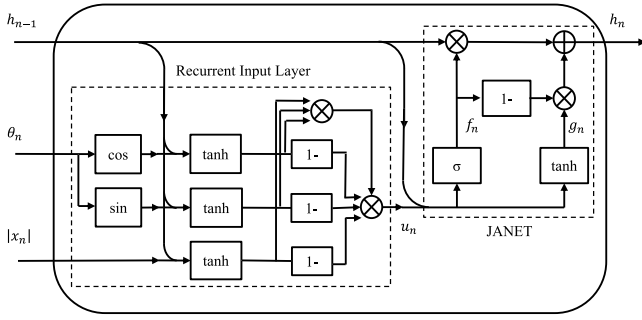
Fig. 1. PG-JANET cell architecture.

how to construct a recurrent DVR model. In Section IV, the complete model structure and the model training process are given. Section V presents experimental validation followed by a conclusion in Section VI.

II. REVIEW OF PA MODELING WITH PHASE-GATED RECURRENT NEURAL NETWORK

With wireless systems demanding further wider signal bandwidth, the consequent longer memory effects have a significant impact on RF PA modeling. To represent the more complex PA behavior dominated by memory effects, recurrent neural networks arise as an alternative solution and have attracted more interests in recent DPD development, as they have proved their proficiency in learning long-term dependencies.

As regards the recurrent neural networks, many network structures addressing different issues can be found in the literature. Instead of using the generic structures, tailoring the best suitable one for modeling complex nonlinear behavior of the RF PAs is more effective. On the one hand, the rightly fit model can accurately catch the physical dynamics of the PA. On the other hand, removing redundant terms is beneficial in reducing the complexity for practical implementation. In our previous work in [22], we carefully examined the relationship between the physical behavior of the PA and the neural network structures and proposed a modified version of recurrent neural networks, called PG-JANET, to model the PA and conduct the DPD.

As depicted in Fig. 1, the recurrent operations in JANET can be described with the functions of forget gate f_n, inner cell g_n, and the output of the network h_n

$$\begin{cases} f_n = \sigma\left(W_f[x_n, h_{n-1}] + b_f\right) \\ g_n = \tanh(W_c[x_n, h_{n-1}] + b_c) \\ h_n = f_n \otimes h_{n-1} + (1 - f_n) \otimes g_n \end{cases} \quad (1)$$

where W_f and W_c are the weights of the layers, and b_f and b_c are the bias, while $[x_n, h_{n-1}]$ stands for concatenation. Moreover, $\otimes$ and σ represent the element-wise multiplication and the sigmoid activation function, respectively. From the results given in [22], we can see that the output function h_n of JANET can be attributed to the PA behavior described via a recurrent feedback structure that can cover long-term memory effects. The JANET network is a single gate simplified model of long short-term memory (LSTM) network, which provides high-accuracy performance with low complexity. Using such

a compact network, thus, lowers model implementation complexity and reduces model running time.

This experimental results in [22] confirmed that processing of current input by associating it with the recurrent information is quite useful to capture the memory effects. In fact, one can rely on this recurrent information to cover longer memory effects, rather than taking the input with multiple memory samples, which is the commonly adopted approach used in the PA modeling. Despite the improvement was made, to linearize ultrawideband PAs, e.g., with over 200-MHz modulated signals, a large number of model coefficients are still required.

III. BUILDING DVR WITH RECURRENT CONNECTIONS

The PG-JANET model showed us how an effective recurrent structure can boost the model capability on capturing long-term memory effects. However, the nonlinear process of the signals still relies on the conventional neural network approach, e.g., linear weighting with simple activation functions, which may not be effective in dealing with the complex nonlinear behavior of the PA. Thus, we believe that integrating a powerful nonlinear processing, which is also well tailored for PA and DPD modeling conditions, into the recurrent learning of the JANET network further enhances the model capability.

In the course of search for such a nonlinear structure, the DVR model [24] emerges as a promising candidate. In this section, we demonstrate how the DVR processing can be integrated with recurrent connections to form a recurrent DVR structure.

A. DVR Model

The DVR model is a modified version of the canonical piecewise linear (CPWL) function [25] for behavioral modeling of RF PAs [24]. Using CPWL, the output of a finite-memory nonlinear digital system is approximated by the following equation:

$$y(n) = \sum_{i=0}^{M} a_i x(n-i) + b + \sum_{k=1}^{K} c_k \left| \sum_{i=0}^{M} a_{ki} x(n-i) - \beta_k \right| \quad (2)$$

where $y(n)$ and $x(n)$ denote the output and input, respectively. $|\cdot|$ stands for the absolute value operation, whereas K is the number of the partition, and β_k becomes the threshold, which decides the boundary of the partition. M represents the memory length, and a_i, b, c_k, and $a_{ki,j}$ can be defined as the model parameters. From (2), we can see that the nonlinear process is achieved using the "absolute" operation. It has been proven that the CPWL can be used to represent a wide range of nonlinear behavior of analog circuits with a high precision [25].

The conventional CPWL function, however, only can deal with real-valued signals. To make it suitable for DPD, e.g., to deal with complex signals and make it linear-in-parameters, in [24], the nonlinear basis function of (2) was modified from

$$\sum_{k=1}^{K} c_k \left| \sum_{i=0}^{M} a_{ki} x(n-i) - \beta_k \right| \quad (3)$$

to

$$\sum_{i=0}^{M} \sum_{k=1}^{K} c_{ki} \left| |\tilde{x}(n-i)| - \beta_k \right| e^{j\theta(n-i)} \qquad (4)$$

where the nonlinear process is conducted on the samples first before linearly combined. The inner $|\cdot|$ is to find the magnitude of the baseband input, whereas the outer $|\cdot|$ represents the absolute value operation. The phase information of the sample is recovered by multiplying the output with $e^{j\theta(n-i)}$. In principle, the DVR performs a vector decomposition of the signal into magnitude and phase and then compares the magnitude with multiple predefined thresholds β_k. Next, the nonlinear operation is performed on the this comparison via the "absolute" value operation as CPWL does. The results are multiplied with the corresponding model coefficients and summed up. Finally, the phase restoration comes at the last stage of the implementation. To improve model accuracy, it can also include additional basis functions. The complete DVR model is described by the following equation:

$$\tilde{y}(n)|_{\text{DVR}} = \sum_{i=0}^{M} \tilde{a}_i \tilde{x}(n-i)$$
$$+ \sum_{i=0}^{M} \sum_{k=1}^{K} \tilde{c}_{ki,1} \left| |\tilde{x}(n-i)| - \beta_k \right| e^{j\theta(n-i)}$$
$$+ \sum_{i=0}^{M} \sum_{k=1}^{K} \tilde{c}_{ki,21} \left| |\tilde{x}(n-i)| - \beta_k \right| e^{j\theta(n-i)} |\tilde{x}(n)|$$
$$+ \cdots \qquad (5)$$

where $\tilde{y}_n$, $\tilde{x}_n$, and θ_n denote the baseband output, input, and the phase of the input, respectively, and $\tilde{a}_i$ and $\tilde{c}_{ki,j}$ are model coefficients.

Different from the conventional polynomial-based models, such as GMP, the DVR model is much more flexible, especially in modeling highly nonlinear systems. However, it is still in a feedforward structure, where the memory length that the model can cover depends on the number of memory samples used in the input. This leads that, in a wideband system, the number of memory samples required can be very large, which, as a result, significantly increases the model complexity.

B. Recurrent DVR Scheme

To address the complexity issue of the model, in this work, we propose to incorporate the recurrent structure presented in the PG-JANET model into the DVR nonlinear process to build the DPD model.

1) Magnitude and Phase Recurrent Filters: In a recurrent network, the basic operation can be expressed as follows:

$$y_n = \tanh(W_x \otimes x_n + W_h \otimes h_{n-1} + b) \qquad (6)$$

where W_x, W_h, and b are weight matrices and bias term, respectively. x_n and y_n denote the input and output of the model at time instant n, whereas h_{n-1} represents the previous hidden information. Also, note that tanh specifies the nonlinear activation function for the operation. To incorporate the recurrent operation, we propose to feed the input signal into a magnitude and a phase recurrent filter, respectively, following
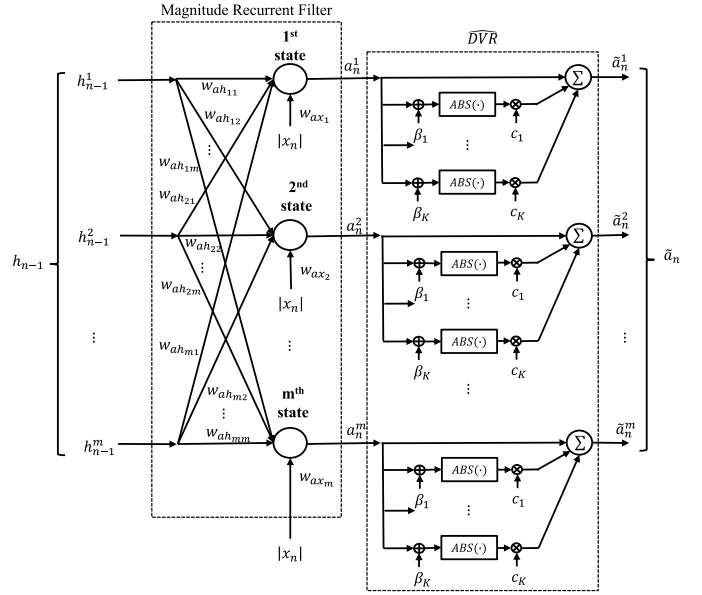

Fig. 2. Nonlinear basis function design.

the decomposition of the baseband data into magnitude and phase. The operation can be described by the following equation:

$$\begin{cases} a_n = W_{ax} \otimes |x_n| + W_{ah} \otimes h_{n-1} \\ \tilde{\theta}_n = W_{p\theta} \otimes \theta_n + W_{ph} \otimes h_{n-1} \end{cases} \qquad (7)$$

where W_{ax} and $W_{p\theta}$ represent the input coefficient matrices, while the weights of the hidden state h_{n-1} are expressed as W_{ah} and W_{ph}, respectively.

The main motivation behind such a filtering of (7) is to combine the input signal with the previous sequence information from the recurrent output, h_{n-1}, by weighing them with the related coefficient matrices. This enables us to associate the magnitude and the phase of complex-valued PA data with a long range of past information in the system.

2) Nonlinear Basis Function Design: By performing the mathematical operation in (7), we arrive at the following expression for each element of a_n, in the magnitude branch:

$$a_n^l = \sum_{i=1}^{m} w_{ah_{li}} h_{n-1}^i + w_{ax_l} |x_n| \qquad (8)$$

where $w_{ah_{li}}$ and w_{ax_l} are the coefficients in the weight matrices of W_{ah} and W_{ax}, respectively. $l = 1, 2, \ldots, m$ is the neuron index, and m determines the number of hidden neurons used.

After obtaining the input, the nonlinear basis function of the DVR model defined in (4) can be directly used to construct the model nonlinear process. For the magnitude branch, to produce the output, a_n^l will be compared with β_k, which is the threshold defined by $\beta_k = k/K$ for $k = 1, 2, \ldots, K$ and then multiplied with the corresponding model coefficients. The resulting nonlinear basis function can be expressed as follows:

$$\sum_{k=1}^{K} c_k \left| \sum_{i=1}^{m} w_{ah_{li}} h_{n-1}^i + w_{ax_l} |x_n| - \beta_k \right|. \qquad (9)$$
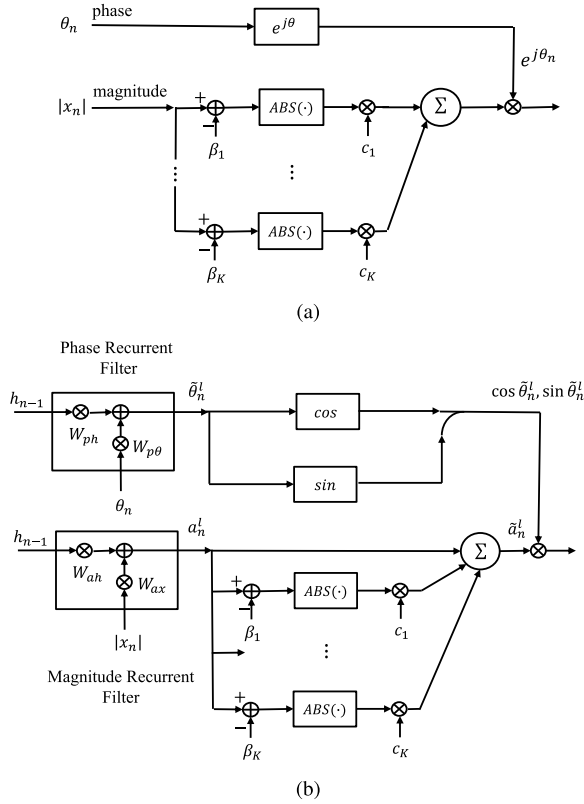
(a)

(b)

Fig. 3. Comparison of vector decomposition and phase restoration between (a) DVR [24] and (b) recurrent DVR scheme.

Note that each of a_n^l defines a different hyperplane partition, as each hidden neuron stores the past information with a different weight. Hence, the previous sequence information differs from each one. Though all the elements are compared with the same threshold β_k and the partitions are weighted by the same coefficient, c_k, each element of the magnitude filter points out a different DVR structure. To clarify the flow of the nonlinear basis function design, Fig. 2 depicts the block diagram of the structure.

3) Complete DVR Scheme: Fig. 3(b) shows how the recurrent DVR scheme works, in comparison with the DVR model itself shown in Fig. 3(a). The input of the DVR model consists of the current input or the time-delayed samples, whereas the recurrent scheme takes the input at that instant and stores it with the hidden information within the neurons of the network to address the previous sequence data. Thus, the recurrent DVR provides a more powerful nonlinear processing in regard of relating the past information to the output. It is because the recurrent state, h_{n-1}, in the basis function of (9) can have infinite memory length theoretically, whereas (4) only can include a limited numbers, e.g., M, of memory samples.

IV. RECURRENT DVR-JANET MODEL

The proposed recurrent DVR unit shown in Fig. 3(b) can be used as a network itself to model the nonlinearity and memory effects of the PA in the course of constructing DPD. However, this might come at the price of vanishing or exploding gradient of the signals, as it lacks the feedback connections through the gating mechanism [26].

To ensure the stability of the recurrent learning and to more accurately address the physical behavior of the PA, this section presents the full model architecture of the proposed DVR-JANET model. To construct DVR-JANET, we combine the DVR scheme in Fig. 3(b), a modified JANET unit, and the final linear output layers.

A. DVR-JANET Model

Considering JANET modification, first, we have simplified the forget gate, f_n, without changing its dynamics. The operation of f_n is simply based on processing the present input and the previous information, and then nonlinearly activating them with the sigmoid function. The interpretation of the sigmoid operation is selectively removing redundant memory states from the model, therefore, focusing on the important ones for the future prediction. At this point, if we consider the usual behavior of PAs, we see that the output signal is mostly dominated by the instantaneous input sample [27]. Building on that, we can release the current input from the forget operation, and therefore, only the hidden information is processed through the forget gate, which lowers the overall complexity.

The other significant difference from the original JANET is that the network has two inner cells, $g_{\cos_n}$ and $g_{\sin_n}$, instead of g_n, which is illustrated in Fig. 1, and can be regarded as the inner memory of the JANET. With the additional inner state, now, we can define separate $W_{c_{\cos}}$ and $W_{c_{\sin}}$ as the weights of each layer. The reason of this modification is based on that the output of DVR scheme consists of two elements: $\tilde{a}_n \cos(\tilde{\theta}_n)$ and $\tilde{a}_n \sin(\tilde{\theta}_n)$. Instead of concatenating these through one inner cell, encoding the cosine and sine components of the phase restoration in separate two inner memories will provide the network a better learning capability for each information.

Thus, using the newly modified forget gate and these two inner memory cells, the network generates two outputs, h_n^I and h_n^Q, in other words, two recurrent states, which is also essentially different from the outcome in the conventional JANET. Both of h_n^I and h_n^Q functions are calculated identically to the output function of JANET defined in (1) using $g_{\cos_n}$ and $g_{\sin_n}$ separately. One can especially consider h_n^I and h_n^Q as the estimated in-phase and the quadrature states of the DVR-JANET recurrent cell, since they are built on the information of $\tilde{a}_n \cos \tilde{\theta}_n$ and $\tilde{a}_n \sin \tilde{\theta}_n$, respectively. The complete model structure is illustrated in Fig. 4, and the operation of DVR-JANET can be described by the following equations:

$$\begin{cases} \tilde{a}_n = \widehat{DVR}\left(W_{ax} \otimes |x_n| + W_{ah} \otimes \left(h_{n-1}^I \oplus h_{n-1}^Q\right)\right) \\ \tilde{\theta}_n = W_{p\theta} \otimes \theta_n + W_{ph} \otimes \left(h_{n-1}^I \oplus h_{n-1}^Q\right) \\ f_n = \sigma\left(W_f\left[\left(h_{n-1}^I \oplus h_{n-1}^Q\right)\right] + b_f\right) \\ g_{\cos_n} = \tanh\left(W_{c_{\cos}}\left[h_{n-1}^I, \tilde{a}_n \otimes \cos \tilde{\theta}_n\right] + b_{g_{\cos}}\right) \\ g_{\sin_n} = \tanh\left(W_{c_{\sin}}\left[h_{n-1}^Q, \tilde{a}_n \otimes \sin \tilde{\theta}_n\right] + b_{g_{\sin}}\right) \\ h_n^I = f_n \otimes h_{n-1}^I + (1 - f_n) \otimes g_{\cos_n} \\ h_n^Q = f_n \otimes h_{n-1}^Q + (1 - f_n) \otimes g_{\sin_n} \\ I_{\text{pred}_n} = W_{o_1} h_n^I \\ Q_{\text{pred}_n} = W_{o_2} h_n^Q. \end{cases} \quad (10)$$
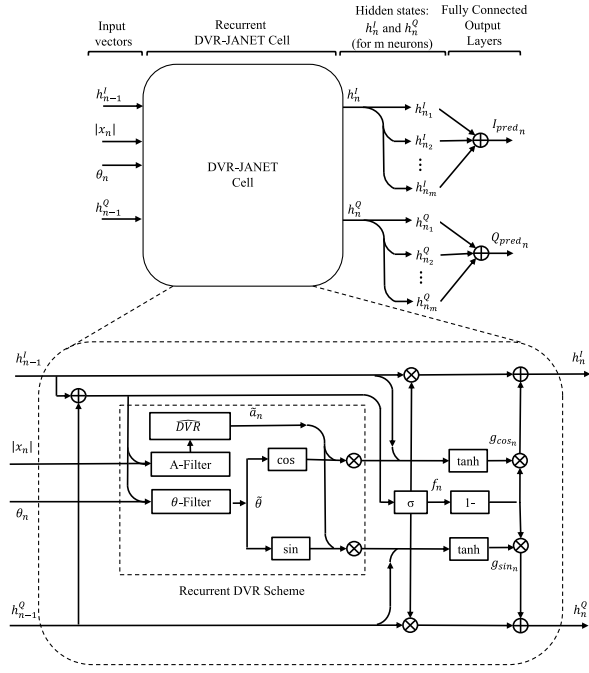
Fig. 4. Folded architecture of the recurrent DVR-JANET model.

As the input layer before the modified JANET, we implemented the recurrent DVR idea discussed in Fig. 3(b) with some alterations. Note that the DVR-JANET cell has two hidden states carrying different past feedback data: h_{n-1}^I and h_{n-1}^Q. They are combined with the element-wise addition operator, as the consequent past sequence information for the magnitude and the phase filters.

With the history information, the recurrent filters perform a linear mapping on the input sequences, $|x_n|$ and θ_n. The output of the magnitude filter continues to be processed with the nonlinear basis function defined in (9), whose operation is represented with $\widehat{DVR}$, and the outcome is noted as $\tilde{a}_n$ in (10) and Fig. 4. Then, the real-valued phase restoration takes place with the outputs coming from $\tilde{a}_n$, $\cos\tilde{\theta}_n$, and $\sin\tilde{\theta}_n$ and results in two components, $\tilde{a}_n\cos\tilde{\theta}_n$ and $\tilde{a}_n\sin\tilde{\theta}_n$. As discussed before, they pass through $g_{\cos_n}$ and $g_{\sin_n}$, each of which only uses its own recurrent state, h_{n-1}^I and h_{n-1}^Q, respectively. On the other hand, note that such as the recurrent filters, the forget gate, f_n, also conducts its operation with the combination of both recurrent connections.

As DVR-JANET unit completes all recurrent functions described in (10) and generates the hidden states, consequently, they are fed to the final linear layers whose coefficient matrices are W_{o_1} and W_{o_2}, to generate the predicted in-phase (I_{pred_n}) and the quadrature (Q_{pred_n}) parts of the complex-valued predistorted signal, i.e., the output of the complete model.

B. Training of DVR-JANET

Considering the acquired behavioral model of the PA, the same model architecture can be exactly used for DPD by swapping the input and output of the PA. This section presents a method for how to train DVR-JANET to define

---

Algorithm 1 DVR-JANET Training Algorithm

Pretraining Stage: ILC Test

Input: x and y
Output: x and $\mathbf{u_{ilc}}$
1: Minimize the error between **x** and **y** iteratively
2: Obtain the dataset of the target DPD for the training

Learning Stage: Neural Network Training

Input: x and $\mathbf{u_{ilc}}$
Output: Model Coefficients
3: Train the network with **x** and $\mathbf{u_{ilc}}$
4: Determine the model coefficients

Adaptive Stage: DLA-based Fine Tuning

Input: The weights of DVR-JANET cell, $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$
Output: Updated Model Coefficients
5: Use the coefficients of the DVR-JANET cell directly
6: Tune the parameters of the fully connected linear layers in a DLA scheme for the final prediction
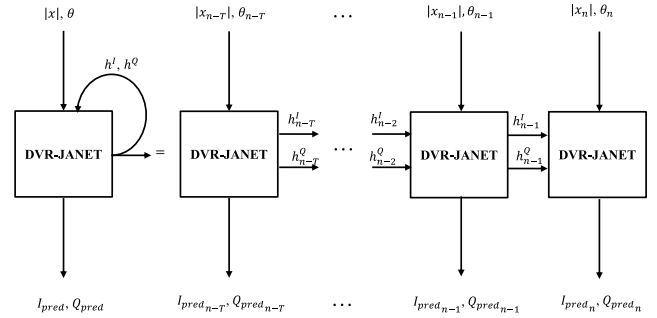

Fig. 5. Unfolded DVR-JANET learning with T time steps.

a DPD model. Entire procedure consists of three stages, as previously described in [22] and also summarized in Algorithm 1.

When deploying a neural network-based DPD, we essentially need a training dataset. To teach the proposed neural network to identify the target DPD, this dataset is obtained from the ILC method [28]. The ILC unit minimizes the error between the input, x, and the output, y, of the PA iteratively updating its output to the controlled system. Therefore, it determines the ideal input u_{ilc} of the PA for the minimum error achieved and provides the training dataset consisting of x and u_{ilc} for the DVR-JANET learning.

After the preparation stage, next step attempts to identify the weights of the DPD model. Fig. 5 summarizes the sequential learning of DVR-JANET by illustrating the unfolded network with T time steps. The folded DVR-JANET in the left-hand side simply shows the input and the output vectors together with the recurrent states of the network. The DVR-JANET block processes the input consisting of the magnitude vector $|x|$ and the phase vector θ by directly feeding the previous hidden information h^I and h^Q back to the model and then predicts the output $y = [I_{\text{pred}}, Q_{\text{pred}}]$.

The recurrent DVR-JANET decides the model parameters over many time steps. We can simplify this learning by unfolding its representation along the input sequence, which
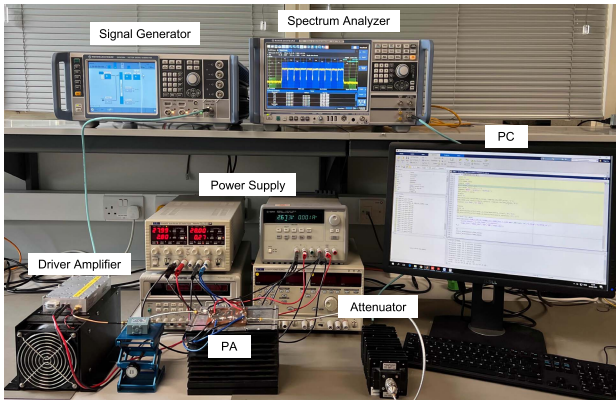
Fig. 6.   Experimental setup.

is illustrated in the right part of Fig. 5. This is a useful conceptual visualization to clarify how the network processes the input samples during the forward pass. At time instant n, for example, the model takes the new input $|x_n|$ and θ_n and uses the hidden state information h_{n-1}^I and h_{n-1}^Q in the memory cells; then, it generates the model output based on (10). After one complete pass of the input sequence, we arrive at predicted multiple time steps of the output vector. It should be especially noted that the DVR-JANET network does not change between the time steps. It is always the same block using the same set of parameters for each time step until the next update of model coefficients.

Finally, we add an adaptive stage to the training called direct learning (DLA)-based fine-tuning. The ultimate goal of this stage is to update the DPD occasionally in case any variation in the input–output, $\hat{x}$–$\hat{y}$, characteristics happens [29]. During this stage, we retain and reuse the coefficients of the recurrent unit, while we fine-tune the weights of the final linear layers before the final model prediction. Considering that the fine-tuning is a pure linear system identification problem, it results in a very low complexity.

V. EXPERIMENTAL VALIDATION

To validate the linearization performance of the proposed model, different experimental tests were conducted, and the results are presented in this section.

A. Experimental Setup and Configurations

Fig. 6 demonstrates the experimental setup including a test computer (PC) running MATLAB and PyTorch software, a vector signal generator (SMW200A) of Rohde & Schwarz controlled via the test computer, a linear driver amplifier, a PA, a -30-dB RF attenuator, and a spectrum analyzer (FSW50) from Rohde & Schwarz. The PA under test is an in-house designed broadband gallium nitride (GaN) Doherty PA operated at 2.80–3.55 GHz with 9.3–11.1-dB gain and 43.0–45.0-dBm saturated power [30]. During the experiment, the input signal is generated in MATLAB running on the PC and then sent to the vector signal generator, which forms the baseband signal and upconverts it to the carrier frequency of 3.2 GHz. After passing the linear driver amplifier, the upconverted signal becomes the input of the PA. The output is attenuated before coming to the spectrum analyzer, which finally downconverts and digitizes the output signal to be saved in the PC.

Throughout the experiment, the proposed model DVR-JANET was compared with our prior work PG-JANET, DVR model itself, and GMP as the conventional model. It is worth to mention that we have compared PG-JANET model with other neural network-based models, particularly the state-of-the-art time-delayed neural network and LSTM-based models, such as the augmented vector-decomposed time-delayed network (AVDTDNN) model [15] and the vector decomposed long short-term memory (VDLSTM) model [17], in [22]. To avoid replication, we did not conduct tests with other models, because they are all worse than the PG-JANET.

As two different wideband scenarios, the measurements were performed with the test signals of 100- and 200-MHz orthogonal frequency-division multiplexing (OFDM) with 7.7-dB peak-to-average power ratio (PAPR). The average output power of the PA was 35.7 dBm. After the ILC test, 160 000 I/Q samples were recorded with the sampling frequencies of 400 and 800 MHz, respectively, for the training of the proposed DVR-JANET and PG-JANET both of which had one recurrent layer. They were trained with the batch size of 40 during 500 epochs and tested with different numbers of hidden neurons. To update network weights, the adaptive moment estimation (ADAM) [31] optimizer was adopted, and to enhance the modeling accuracy, a decaying learning rate method is followed, which decreases the learning rate by 10 after learning curve saturates. For DVR and GMP, the same number of I/Q samples was used for the model extraction, and their configurations also differ in various tests.

As the performance comparison metrics, we presented normalized mean square error (NMSE) and adjacent channel power ratio (ACPR). To compare the model complexity, the number of real-valued parameters was considered as the main criteria, which suggests that a complex-valued parameter of DVR and GMP was counted as two real-valued free parameters during the comparison.

B. Experimental Results

1) Test Results of 100-MHz OFDM Signal: The first wideband experimental validation was performed with five-carrier test signal of 100 MHz. This measurement aims to carry out an extensive search to understand how the models perform with different numbers of model parameters and to present a broad comparison by taking various configurations of them into consideration eventually.

In this test, the number of hidden states was swept from 8 to 14 for DVR-JANET and PG-JANET, and DVR and GMP were specifically configured, so that they can attain their best performance for each evaluation point. Figs. 7 and 8 depict how NMSE and ACPR of the models change by increasing the number of model parameters. It is worth mentioning that the same number of model parameters does not necessarily means exactly the same model complexity. Depending on the model structure and hardware implementations, different models may
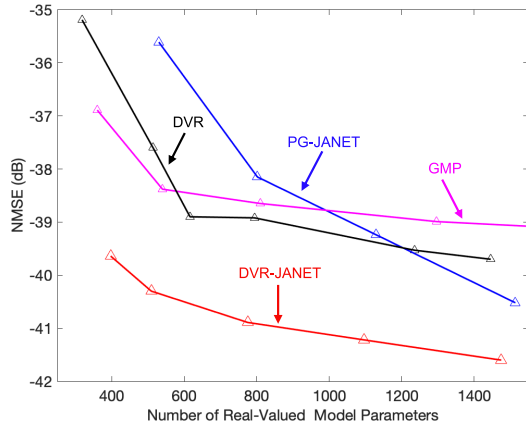
Fig. 7.   NMSE (dB) of the compared models with different numbers of real-valued model parameters for 100-MHz OFDM test.
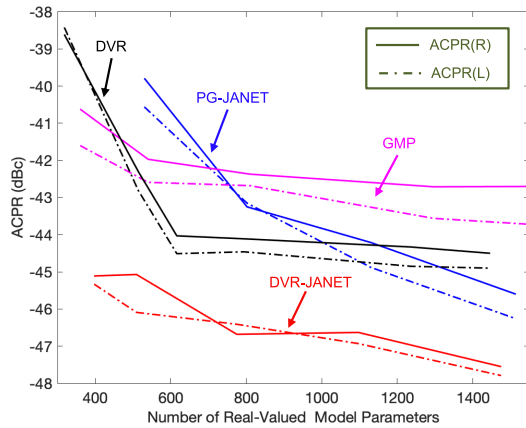

Fig. 8.   ACPR (dBc) of the compared models with different numbers of real-valued model parameters for 100-MHz OFDM test.

TABLE I
PERFORMANCE COMPARISON OF THE MODELS
FOR THE 100-MHz OFDM SIGNAL

Model Type	NMSE (dB)	ACPR (dBc) (lower/upper)	Number of Parameters (real-valued)
Without DPD	-26.43	-36.23/-29.96	N/A
GMP	-38.25	-42.75/-42.08	540
DVR	-37.81	-43.48/-42.56	538
PG-JANET	-35.60	-40.56/-39.80	530
DVR-JANET	-40.27	-46.10/-45.10	509
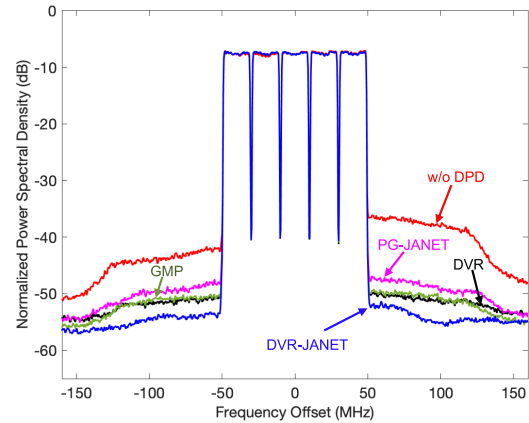
Fig. 9.   Power spectral density comparison between the models for 100-MHz OFDM.

lead to different resource usage. But, more or less, the number of model parameters gives an indication of model complexity. For this test, any of the models can be adopted with their own specific configuration for a linearization task considering the indicators of the model performance in Figs. 7 and 8.

Nevertheless, among the compared models, GMP becomes the most suffering one from the problem of the restricted capability despite of the increasing model parameters. Considering its quickly saturated performance, it can be deduced that there is almost no room for further improvement. DVR and PG-JANET, however, allow more flexibility to accomplish better modeling performance with higher number of model coefficients. Here, we should remark that PG-JANET still can provide much better NMSE and ACPR than the best result of DVR after reaching a particular complexity level.

Ultimately, DVR-JANET appears to have a great advantage over the compared models in each comparison metric. It not only results in the most successful linearization all the time, but also it distinctively provides a much better performance even with a relatively small number of model parameters. For further discussion, to closely investigate the modeling capability of the models in a similar complexity, we can limit the number of model parameters within a particular range

and evaluate the model performance. This enables us a fair comparison between the models under a complexity budget.

For this purpose, we can probe into the comparison when each model has ≈ 500 real-valued parameters, and the DVR-JANET achieves -45 dBc in ACPR for the 100-MHz long-term evolution (LTE) signal. The test results can be found in Table I, which demonstrates the NMSE, ACPR, and the total number of the model parameters. In this specific case, PG-JANET and DVR-JANET were tested with eight hidden states and $K = 3$ for DVR-JANET. We used the GMP model with only lagging terms with a memory depth of 5, the polynomial order of 5, and a cross term length of 8. Finally, the DVR model had 12 partitions with a memory length of 5, including the linear terms, the first-order basis, the second-order type-1 terms, the second-order type-2 terms, the second-order type-3 terms, and the DDR-1 terms [24].

Table I and Fig. 9 confirm that the proposed DVR-JANET exhibits a distinguished competence, when a less complex model is strictly needed. It explicitly brings about a better performance, which is at least 2 dB better in both NMSE and ACPR when compared with DVR and GMP. In case we compare PG-JANET and DVR-JANET, almost 5-dB difference in both of the performance metrics can be observed. Even though PG-JANET is a quite powerful model that can attain much better results than conventional DPDs, it may no longer have the capability to model the PA effectively if the sources are limited. The AM/AM and AM/PM characteristics of the PA can be seen in Fig. 10, showing the behavior both without DPD and with DPD based on DVR-JANET.
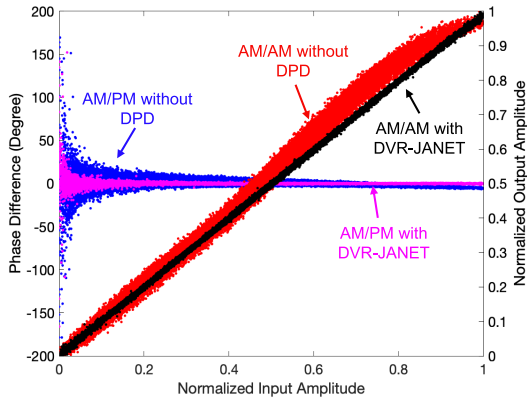
Fig. 10. AM/AM and AM/PM characteristics with and without DPD for the 100-MHz OFDM signal.
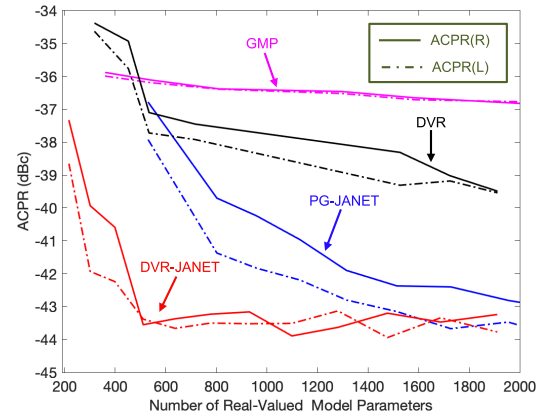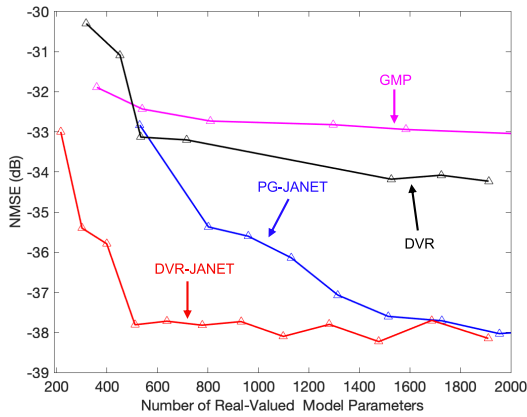

Fig. 11. NMSE (dB) of the compared models with different number of real-valued model parameters for 200-MHz OFDM test.

2) Test Results of 200-MHz OFDM Signal: To validate the performance of the proposed model in an even wider bandwidth, the second test was conducted with ten-carrier 200-MHz OFDM signal. This measurement basically examines how the model performs when the linearization problem becomes more challenging due to the higher distortion induced by the PA.

Similar to the first measurement, this experimental work intends to present a comprehensive comparison between the models under several configurations of them, which leads to different number of model parameters. In this regard, Figs. 11 and 12 provide not only the results of various measurements in terms of NMSE and ACPR, but also an important insight regarding their modeling capabilities. To obtain these results, the number of hidden states of DVR-JANET and PG-JANET was swept from 6 to 17 and from 8 to 18, respectively. Besides, DVR and GMP were tuned to their best performance again as discussed previously.

In the presence of stronger distortion, the conventional models demonstrate a severe lack of an effective DPD modeling, which can be observed from Figs. 11 and 12. Apparently, the higher model complexity does not allow them to improve their modeling accuracy. On the other hand, PG-JANET and DVR-JANET obviously can reach a particular linearization performance, which emphasizes the importance of a flexible

Fig. 12. ACPR (dBc) of the compared models with different numbers of real-valued model parameters for 200-MHz OFDM test.

TABLE II
PERFORMANCE COMPARISON OF THE MODELS
FOR THE 200-MHz OFDM SIGNAL

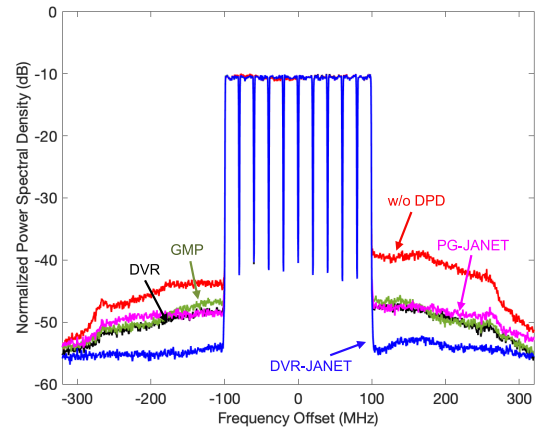| Model Type	NMSE (dB)	ACPR (dBc) (lower/upper)	Number of Parameters (real-valued)
Without DPD	-24.82	-31.94/-28.06	N/A
GMP	-32.43	-36.19/-35.73	540
DVR	-33.14	-37.72/-36.62	534
PG-JANET	-32.83	-37.92/-36.78	530
DVR-JANET	-37.82	-43.40/-43.55	511

Fig. 13. Power spectral density comparison between the models for 200-MHz OFDM.

and effective neural network modeling. From Figs. 11 and 12, we can conclude that the best possible performance of DVR-JANET and PG-JANET converges toward nearly the same level. However, DVR-JANET obviously arrives at the targeted performance with a much less complex model configuration, which has nearly one-third of the PG-JANET model parameters. Based on this, we can conclude that the DVR-JANET shows a great improvement on the learning ability of PG-JANET.

Similarly, the performance of the compared models is presented in Table II when they have nearly 500 real-valued parameters. The output spectrum, and the AM/AM and AM/PM
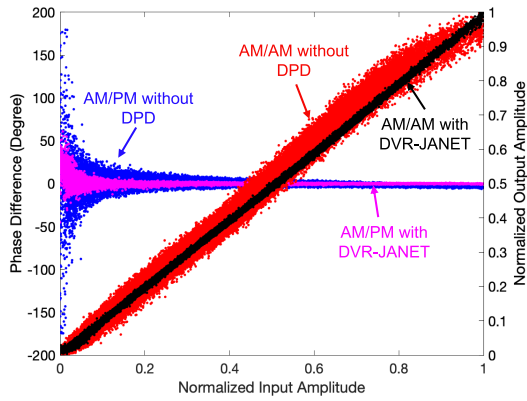
Fig. 14.   AM/AM and AM/PM characteristics with and without DPD for 200-MHz OFDM signal.

plots are also depicted in Figs. 13 and 14, respectively. For this test, PG-JANET and DVR-JANET were configured with eight hidden states and $K = 5$ for DVR-JANET. While GMP was tested with the same configuration employed in 100-MHz band, the DVR model configuration changed to the partition number of 10 and the memory length of 10, including the linear terms, the first-order basis, the second-order type-1 terms, and the second-order type-2 terms.

In this case, both Table II and Fig. 13 indicate that only DVR-JANET can provide a competent compensation of the nonlinear distortion within this particular complexity range, while the others result in a poor linearization performance. This is because, with wideband signals, the nonlinear behavior of the PA becomes much more complex, e.g., longer memory effects would occur, and the interaction between memory samples becomes more prominent. The conventional feedforward models would not be able to handle the long-term memory effects, while PG-JANET cannot model the nonlinear interaction accurately. In this case, combining recurrent network with high precision piecewise DVR function certainly makes sense, which verifies the superiority of DVR-JANET over the other compared models.

VI. Conclusion

To reduce relatively higher complexity of wideband DPDs, we propose a novel DVR-based recurrent neural network model for RF PA linearization in this article. Building a novel recurrent DVR structure and integrating it into a gated recurrent network, we aim to enhance the learning capability of our prior work PG-JANET, therefore, to achieve the same performance with a smaller number of model coefficients. Experimental tests have demonstrated that the proposed model not only improves the modeling accuracy, but also lowers the model complexity significantly. Based on its high linearization performance and distinct advantage in the complexity aspect, we believe that the proposed DVR-JANET model provides a promising solution to future challenges of wideband DPD applications.

References

[1] Z. Popovic, "Amping up the PA for 5G: Efficient GaN power amplifiers with dynamic supplies," *IEEE Microw. Mag.*, vol. 18, no. 3, pp. 137–149, May 2017.

[2] L. Guan and A. Zhu, "Green communications: Digital predistortion for wideband RF power amplifiers," *IEEE Microw. Mag.*, vol. 15, no. 7, pp. 84–99, Nov. 2014.

[3] F.-L. Luo, *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[4] L. Ding *et al.*, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[5] 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.

[6] O. Hammi, F. M. Ghannouchi, and B. Vassilakis, "A compact envelope-memory polynomial for RF transmitters modeling with application to baseband and RF-digital predistortion," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 5, pp. 359–361, May 2008.

[7] 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.

[8] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Open-loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 7, pp. 1524–1534, Jul. 2008.

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

[10] 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. 2009.

[11] M. Rawat and F. M. Ghannouchi, "Distributed spatiotemporal neural network for nonlinear dynamic transmitter modeling and adaptive digital predistortion," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 3, pp. 595–608, Mar. 2012.

[12] E. G. Lima, T. R. Cunha, and J. C. Pedro, "A physically meaningful neural network behavioral model for wireless transmitters exhibiting PM–AM/PM–PM distortions," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3512–3521, Dec. 2011.

[13] 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 Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan. 2019.

[14] R. G. Sáez and N. M. Marqués, "RF power amplifier linearization in professional mobile radio communications using artificial neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3060–3070, Apr. 2019.

[15] 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. 91559–91568, 2019.

[16] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.

[17] 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. 63780–63789, 2020.

[18] G. Xu, H. Yu, C. Hua, and T. Liu, "Chebyshev polynomial-LSTM model for 5G millimeter-wave power amplifier linearization," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 6, pp. 611–614, Jun. 2022.

[19] G. Li, Y. Zhang, H. Li, W. Qiao, and F. Liu, "Instant gated recurrent neural network behavioral model for digital predistortion of RF power amplifiers," *IEEE Access*, vol. 8, pp. 67474–67483, 2020.

[20] M. Tanio, N. Ishii, and N. Kamiya, "Efficient digital predistortion using sparse neural network," *IEEE Access*, vol. 8, pp. 117841–117852, 2020.

[21] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[22] T. Kobal, Y. Li, X. Wang, and A. Zhu, "Digital predistortion of RF power amplifiers with phase-gated recurrent neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 6, pp. 3291–3299, Jun. 2022.

[23] J. van der Westhuizen and J. Lasenby, "The unreasonable effectiveness of the forget gate," 2018, *arXiv:1804.04849*.

[24] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 737–744, Feb. 2015.

[25] L. O. Chua and A.-C. Deng, "Canonical piecewise-linear representation," *IEEE Trans. Circuits Syst.*, vol. 35, no. 1, pp. 101–111, Jan. 1988.

[26] C. Tallec and Y. Ollivier, "Can recurrent neural networks warp time?" 2018, *arXiv:1804.11188*.

[27] Y. Li, W. Cao, and A. Zhu, "Instantaneous sample indexed magnitude-selective affine function-based behavioral model for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 5000–5010, Nov. 2018.

[28] 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.

[29] Y. Li, X. Wang, and A. Zhu, "Complexity-reduced model adaptation for digital predistortion of RF power amplifiers with pretraining-based feature extraction," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 3, pp. 1780–1790, Mar. 2021.

[30] M. Li, J. Pang, Y. Li, and A. Zhu, "Bandwidth enhancement of Doherty power amplifier using modified load modulation network," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 6, pp. 1824–1834, Jun. 2020.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Repr. (ICLR)*, May 2015, pp. 1–15.

Tugce Kobal (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical and electronics engineering from Middle East Technical University (METU), Ankara, Turkey, in 2014 and 2018, respectively. She is currently pursuing the Ph.D. degree at the RF and Microwave Research Group, University College Dublin (UCD), Dublin, Ireland.

Her current research interests include behavioral modeling of RF power amplifiers and digital predistortion based on neural networks.
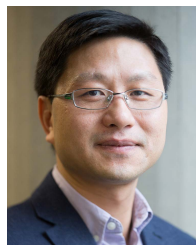
Anding Zhu (Fellow, IEEE) received the Ph.D. degree in electronic engineering from the University College Dublin (UCD), Dublin, Ireland, in 2004.

He is currently a Professor with the School of Electrical and Electronic Engineering, UCD. His research interests include high-frequency nonlinear system modeling and device characterization techniques, high-efficiency power amplifier design, wireless transmitter architectures, digital signal processing, and nonlinear system identification algorithms.

Prof. Zhu is an Elected Member of the IEEE Microwave Theory and Technology Society (MTT-S) Administrative Committee (AdCom) and a member of the IEEE Future Directions Committee. He was a recipient of the 2021 IEEE MTT-S Microwave Prize. He served as the Secretary of the IEEE MTT-S AdCom in 2018 and the Chair of IEEE MTT-S Microwave High-Power Techniques Committee (TC-12) in 2020 and 2021. He is also the Chair of the Electronic Information Committee and the Vice Chair of the Marketing and Communications Committee. He is a Track Editor of IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES and an Associate Editor of *IEEE Microwave Magazine*.