Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters

Dongming Wang, Mohsin Aziz, Mohamed Helaoui, *Member, IEEE*, and Fadhel M. Ghannouchi, *Fellow, IEEE*

***Abstract*— A digital predistorter, modeled by an augmented real-valued time-delay neural network (ARVTDNN), has been proposed and found suitable to mitigate the nonlinear distortions of the power amplifier (PA) along with modulator imperfections for a wideband direct-conversion transmitter. The input signal of the proposed ARVTDNN consists of Cartesian in-phase and quadrature phase (I/Q) components, as well as envelope-dependent terms. Theoretical analysis shows that the proposed model is able to produce a richer basis function containing both the desired odd- and even-order terms, resulting in improved modeling capability and distortion mitigation. Its actual performance has been validated through extensive simulations and experiments. The results show that the compensation and hardware impairment mitigation capabilities of the ARVTDNN are superior to the existing state-of-the-art real-valued focused time-delay neural network (RVFTDNN) by 3–4 dB for the adjacent channel power ratio and by 2–3 dB in terms of the normalized mean square error. Other important features of the proposed model are its reduced complexity, in terms of the number of parameters and floating-point operations, and its improved numerical stability compared to the RVFTDNN model.**

***Index Terms*— Augmented real-valued time-delay neural network (ARVTDNN), dc offset, in-phase and quadrature phase (I/Q) imbalance, neural network (NN), nonlinear distortions, power amplifier (PA), predistortion.**

I. Introduction

A TRANSMITTER modulates a baseband signal to a predefined carrier frequency while amplifying it to a sufficiently high power level to enable its transmission by a wireless medium to a far-off receiver. Modulation is achieved by using a quadrature modulator, which ideally maintains a 90° phase difference between the in-phase and quadrature phase (I and Q) components, while also keeping the same gain and phase between the two branches. However, there is a gain and phase mismatch introduced by the nonideal behavior of the modulator, resulting in I/Q imbalance which degrades the performance of the system.

Manuscript received May 18, 2016; revised December 3, 2016, April 7, 2017, and October 6, 2017; accepted May 9, 2018. Date of publication June 12, 2018; date of current version December 19, 2018. *(Corresponding author: Dongming Wang.)*

The authors are with the iRadio Laboratory, Department of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada, (e-mail: dongmwan@ucalgary.ca; azizm@ucalgary.ca; fghannou@ucalgary.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2018.2838039

Amplification is achieved by using a power amplifier (PA) [1], [2], which is a highly nonlinear and energy-consuming device. PA nonlinearity leads to spectrum regrowth; in which case, the signal may fail to meet the spectral mask requirements and the bit error rate (BER) deteriorates. Moreover, some modern modulation techniques, such as orthogonal frequency division multiplexing (OFDM), result in an increased peak-to-average power ratio (PAPR). A signal with a higher PAPR is more sensitive to PA nonlinearity at the same average power level, leading to an increased nonlinear distortion.

Over the past two decades, numerous linearization techniques combating PA nonlinear distortions have been proposed in an effort to address the issue of spectral regrowth. These techniques include feedback [1], feedforward [2], [3], and digital predistortion (DPD) [4]–[7]. PA nonlinearity and modulator imperfections, as detailed in [8] and [9], further deteriorate the performance of the wireless transmitting system and degrade the quality of the output RF signal.

Various methods have been put forward to address the issues of I/Q imbalance, dc offset, and PA nonlinear distortions [8], [10], [11]. However, many of these methods are two-step processes [10], [11], in which I/Q imbalance and local oscillator (LO) leakage are considered first, followed by PA linearization.

A real-valued focused time-delay neural network (RVFTDNN) model [8], which involves one-step mitigation procedure, has been proposed to simplify the compensation process. As an neural network (NN) model, the RVFTDNN not only exhibits one-step compensation capability, but also presents reasonable performance in the presence of various distortions. Due to their strong adaptive nature and approximation capability, NNs are very attractive for the behavioral modeling of PAs. Techniques, such as a fuzzy logic NN [12], a feedforward NN [13], and a radial basis NN [14], have recently been suggested for constructing the nonlinear behavior of a device under test (DUT).

This paper proposes using an augmented real-valued time-delay NN (ARVTDNN) model for compensation of the direct-conversion transmitter's imperfections. Although the RVFTDNN performs reasonably well for inverse modeling of the transmitter, we have shown that this model, in its original form, has limitations in terms of generating the desired even-order intermodulation products [13].

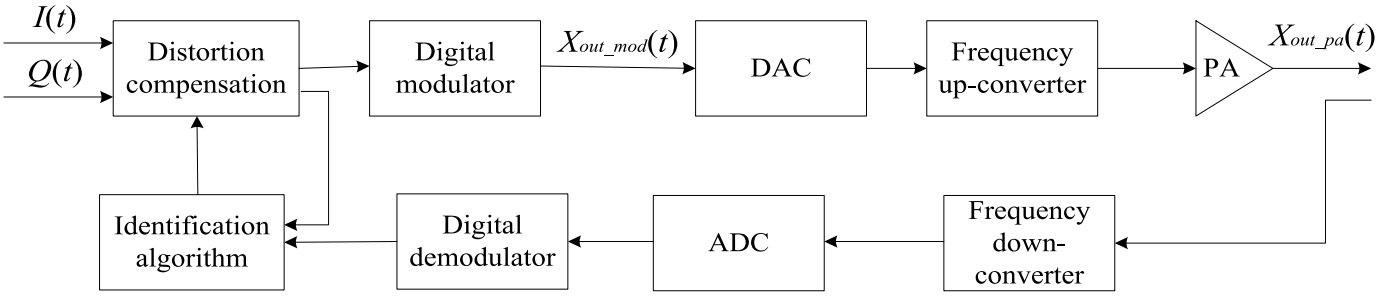

Fig. 1. Block diagram of direct-conversion transmitter.

In order to evaluate the performance of the proposed ARVTDNN, an extensive comparative study between the ARVFTDNN and other existing models, including the RVFTDNN, was performed using experiments and simulations. The results, quantified in terms of various figures of merit, demonstrate the improved performance of the proposed ARVTDNN compared to existing state-of-the-art models.

The following section briefly reviews direct-conversion transmitter's imperfections and explains the structure of NNs. The structure and training process of the proposed ARVTDNN are elaborated in Section III. Section III also provides the theoretical justification of the proposed ARVTDNN while comparing it with the RVFTDNN. In Section IV, the measurement setup and transmitter characteristics under different scenarios are presented. Section V describes the measurement results and comparisons of the proposed model to various other models. Conclusion followed in Section VI.

II. Direct-Conversion Transmitters' Impairments and Neural Networks—A Review

A. Impairments in Direct-Conversion Transmitters

There are various impairments that affect the quality of the transmission signal due to the imperfections of communication devices used in the network [8]. A typical direct-conversion transmitter is shown in Fig. 1. In a transmitter, the dc offset, I/Q imbalance, and PA nonlinearity have enormous effects on the quality of the transmitted signal. The dc offset and I/Q imbalance caused by the modulator can generate distortions, resulting in deterioration of the transmitter's performance. The impaired signal at the output of the modulator can be expressed as [8]

$$X_{\text{out\_mod}}(t) = (I(t) + C_i)\cos(w_c t) - \alpha(Q(t) + C_q)\sin(w_c t + \phi) \quad (1)$$

where $I(t)$ and $Q(t)$ represent the I/Q components of the baseband input signal, C_i and C_q are the dc offsets of the I and Q channels, α and ϕ represent the amplitude and phase imbalances between the I and Q channels, respectively, and w_c is the angular frequency [8].

The PA in the transmitter continuously amplifies the imperfections of the modulator, but also generates significant nonlinear distortions. These distortions severely damage the quality of the signal. The output signal of the PA can be written as [8]

$$X_{\text{out\_pa}}(t) = G \cdot f_{\text{PA}}(X_{\text{out\_mod}}(t)) \quad (2)$$

where $f_{\text{PA}}(\cdot)$ represents the nonlinear characteristics of the PA, and G is PA's small-signal gain (SSG).

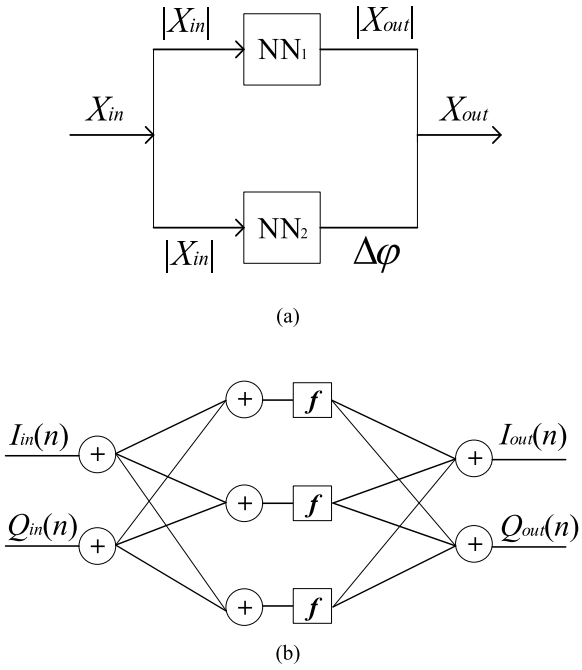

Fig. 2. Conventional NN topologies. (a) Polar topology. (b) Cartesian topology.

B. Neural Networks for Behavioral Modeling of Transmitter Characteristics

As widely shown in the literature, NNs can provide good modeling performance for PAs and reasonable mitigation capability for PA nonlinear distortions. The single-input-single-output feedforward model [8], [15] is a basic structure of NNs that can be used to compensate for PA distortions. However, this architecture introduces complex-valued weights and activation functions, which result in cumbersome calculations.

A polar feedforward NN, as shown in Fig. 2(a), was introduced as a way to extract the amplitude and phase responses separately by employing two NNs [16]. However, the two NN branches in this design usually cannot converge at the same

time, resulting in overtraining or undertraining, as described in [15]. In order to address this issue, a Cartesian NN or real-valued feedforward NN model, as shown in Fig. 2(b), was introduced [8]. This model uses the I and Q signals for the purpose of modeling the transmitter's inverse characteristics and resolves the simultaneous convergence issue. However, as suggested in [8], this topology may not behave satisfactorily when strong memory effects are considered. These effects become, particularly, dominant in wideband signals and need more consideration.

The proposed ARVTDNN model is a hybrid of the Cartesian and polar architectures presented in Fig. 2 and considers both the I and Q components along with the amplitudes of the input signal and its nonlinear versions. The importance of augmenting these envelope terms is detailed in Section III. In addition, the proposed model also includes the memory effects present in the system and is, therefore, quite suitable for the modeling of dynamic systems with large signal bandwidths. The proposed model offers better modeling performance and reduced complexity compared to the existing methods, as demonstrated in this paper.

III. Augmented Real-Valued Focused Time-Delay Neural Network

DPD is an important technique to mitigate transmitters' nonlinear distortions. Among the DPD methods, Hammerstein-based predistortion is a popular technique for the mitigation of PA's nonlinearity. Augmented versions of the Hammerstein model and radial basis function NNs have been presented in [21] and [22], respectively, and provide better modeling and mitigation capabilities. This forms the motivation behind the proposed methodology.

The structure of the ARVTDNN is shown in Fig. 3. It can be observed that the input signal contains the current and past samples of the Cartesian I/Q components along with the envelope-dependent terms. The corresponding input signal vector is represented as

$$\begin{aligned} x(n) = [&I_{\text{in}}(n), I_{\text{in}}(n-1), \ldots, I_{\text{in}}(n-M), \\ &Q_{\text{in}}(n), Q_{\text{in}}(n-1), \ldots, Q_{\text{in}}(n-M), \\ &|X_{\text{in}}(n)|, |X_{\text{in}}(n-1)|, \ldots, |X_{\text{in}}(n-M)|, \\ &|X_{\text{in}}(n)|^2, |X_{\text{in}}(n-1)|^2, \ldots, |X_{\text{in}}(n-M)|^2, \\ &|X_{\text{in}}(n)|^3, |X_{\text{in}}(n-1)|^3, \ldots, |X_{\text{in}}(n-M)|^3] \end{aligned} \quad (3)$$

where $I_{\text{in}}(n)$ and $Q_{\text{in}}(n)$ denote the in-phase and quadrature components of the current samples; $I_{\text{in}}(n-k)$ and $Q_{\text{in}}(n-k)$, $k = 1, 2, \cdots, M$, represent the in-phase and quadrature components of past samples; $|X_{\text{in}}(n)|$ is the amplitude of the current samples; and, $|X_{\text{in}}(n-k)|$ is the amplitude of the past samples. The memory depth M of the input baseband signal is determined by the optimization process [17] and may vary depending on the memory effects exhibited by the system.

A. Neural Network (NN) Training, Validation, and Testing

NNs usually consist of an input layer, hidden layers, and an output layer. The number of hidden layers and their corresponding neurons are commonly determined during training of the network or by using empirical methods [7], [17]. In this paper, one hidden layer was selected based on an empirical method, and the optimal number of neurons in the hidden layer was determined during the training of the network. The number of neurons in the hidden layer was set to be in the range of 7–100. The number of neurons in the input and output layers was determined by the input and output signal vectors and the memory taps required to construct the model.

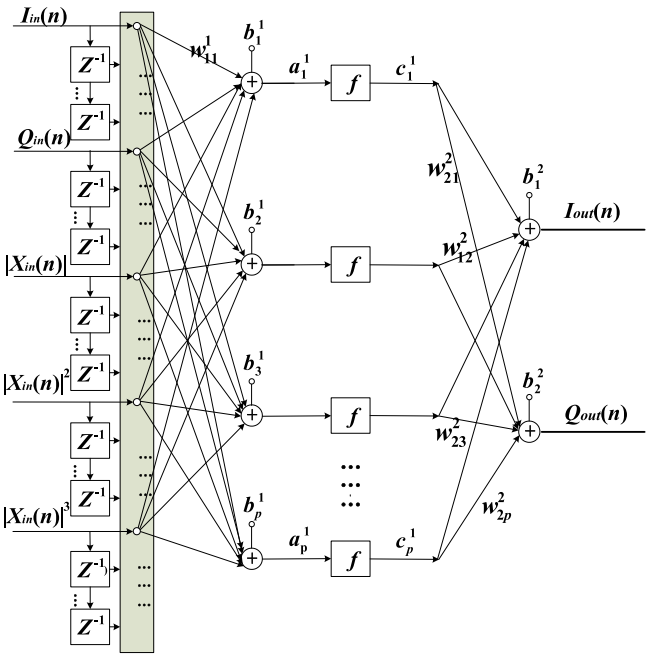

Fig. 3. Block diagram of ARVTDNN.

TABLE I
Models' Performance With Different Activation Functions

Activation functions	NMSE (dB)
logsig	-35.66
purelin	-23.87
satlin	-33.55
poslin	-34.28
satlins	-34.74
tansig	-37.93

A transfer/activation function was used in the hidden layer to achieve the desired nonlinear modeling and mitigation performance. Different transfer functions can be applied, depending on the applications or type of models being considered as well as the dynamic range of the data. Table I elaborates the choice of the activation function.

Another important consideration when using NNs is the length of the data required for the modeling process. In order to save modeling time, the optimal data length has to be determined. The optimal data length can be ascertained by an empirical method or systematically changing the data length used while training the network. In this paper, 7000 samples of the input and output signals were used for training, validation, and testing purposes.

The validation process is used to prevent overtraining, while testing is used to evaluate modeling performance with each iteration using the mean-squared error (MSE) algorithm. 60% of the modeling data were used for the purpose of training, 35% for the validation process, and the remaining 5% in the testing phase. In addition, 4,000 independent sample pairs were used to evaluate the performance of the NN during the last iteration using the normalized MSE (NMSE) [13] as a figure of merit. The output of each neuron in Fig. 3 is equal to the bias plus the sum of the products of the input signals and corresponding weights, which is expressed as [13]

$$a_i^l(n) = \sum_{j=1}^{p} w_{\mathrm{ij}}^l x_j^{l-1}(n) + b_i^l \tag{4}$$

where i denotes the ith neuron of the current layer, j denotes the jth neuron of the previous layer, l represents the lth layer, w_{ij}^l is the synaptic weight connecting the jth neuron of the $(l-1)$th layer to the ith neuron at the lth layer, x_j^{l-1} is the input signal to the jth neuron from the previous $(l-1)$th layer, and b_i^l is the bias of the ith neuron at the lth layer.

The output of any layer is then obtained using the following activation function [13]:

$$c_i^l = f\left(a_i^l(n)\right) \tag{5}$$

where $f(\cdot)$ is the activation function.

Commonly used activation functions namely the log sigmoid (*logsig*), linear transfer function (*purelin*), saturating linear transfer function (*satlin*), symmetric saturating linear transfer function (*satlins*), positive linear transfer function (*poslin*), and hyperbolic tangent sigmoid (*tansig*) were used to evaluate the modeling performance of the NN. These were compared to pick the one with the best performance.

The results are summarized in Table I, where it can be seen that the hyperbolic tangent sigmoid [17] had the best approximation capability for the problem at hand. This function is defined as

$$\tanh(x) = \frac{\exp(2x) - 1}{\exp(2x) + 1} \tag{6}$$

for a given x. Finally, cost function is calculated in batch mode during the forward pass [8], [17]. The cost function is represented as

$$E = \frac{1}{2N} \sum_{n=1}^{N} \left[(I_{\mathrm{est}}(n) - I_{\mathrm{mea}}(n))^2 + (Q_{\mathrm{est}}(n) - Q_{\mathrm{mea}}(n))^2 \right] \tag{7}$$

where $I_{\mathrm{est}}(n)$ and $Q_{\mathrm{est}}(n)$ are the in-phase and quadrature components of the normalized simulated/modeled output signal, while $I_{\mathrm{mea}}(n)$ and $Q_{\mathrm{mea}}(n)$ are the in-phase and quadrature components of the measured/desired output signal, respectively.

After the cost function of the NN is obtained during one epoch/iteration, the values of the synaptic weights and biases are updated using the Levenberg–Marquardt algorithm, a well-known type of the back-propagation algorithm that has been shown to perform exceptionally well [7], [17]. The cost function is then calculated again using the updated synaptic weights and biases during the next epoch/iteration. The hidden layer's synaptic weights are set in the range of −0.8–0.8 [8], in order to avoid extreme values of the activation function.

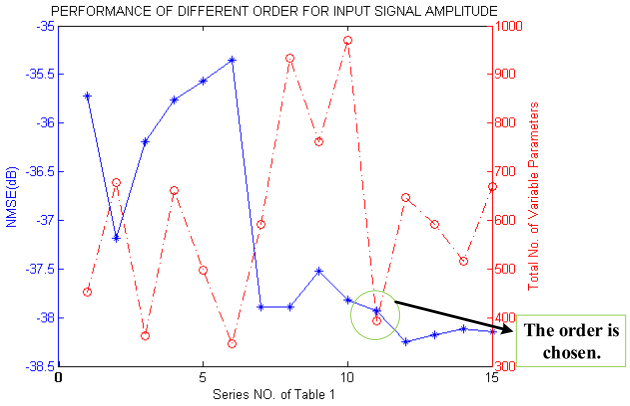

Fig. 4. Performance of NN model for different combinations of amplitude components provided in Table II.

The training process, in terms of the number of iterations required for convergence, is terminated once the NN satisfies the desired modeling performance or reaches the preset maximum epoch/iteration or overtraining [7]. The desired modeling performance in terms of MSE and maximum epochs are set to be 10^{-6} and 100, respectively [17].

B. Selection of the Injected Envelope-Dependent Components

The existing RVFTDNN model considers only I and Q components for modeling purposes. In the proposed model, the envelope-dependent terms of the input signal are injected into the NN along with I and Q components. This modification leads to a richer basis set, capable of better modeling performance.

The choice of the envelope-dependent terms is important to obtain better modeling performance. For this purpose, the impact of different envelope-dependent components on the performance and complexity of the system is listed in Table II. Furthermore, Fig. 4 shows the performance due to these envelope-dependent components in terms of NMSE and complexity. A measure of the performance of the proposed methodology is the NMSE which is defined as [7]

$$\mathrm{NMSE}_{\mathrm{dB}} = 10 \times \log_{10} \frac{\frac{1}{N} \sum_{j=1}^{N} |y_{\mathrm{est}}(n) - y_{\mathrm{mea}}(n)|^2}{\frac{1}{N} \sum_{j=1}^{N} |y_{\mathrm{mea}}(n)|^2} \tag{8}$$

where N is the length of the tested data, and $y_{\mathrm{est}}(n)$ and $y_{\mathrm{mea}}(n)$ are the estimated and measured data, respectively. Complexity is another important parameter that needs to be considered while constructing the model. It is measured in terms of the number of coefficients required for modeling the desired output. The number of coefficients required to compute the model is

$$N_0 = ((M+1) \times I) \times H + H \times O + (H + O) \tag{9}$$

TABLE II
PERFORMANCE OF DIFFERENT ORDERS FOR INPUT SIGNAL AMPLITUDE

Series No.	Amplitude Terms (in addition to I and Q terms)	Optimal No. of neurons in hidden layer	Total No. of variables	NMSE (dB)
Activation function: Hyperbolic Tangent Sigmoid				
1	No amp. Terms (RVFTDNN)	41	453	-35.72
2	$\|X_{in}(n)\|$	45	677	-37.18
3	$\|X_{in}(n)\|^2$	24	362	-36.19
4	$\|X_{in}(n)\|^3$	44	662	-35.76
5	$\|X_{in}(n)\|^4$	33	497	-35.57
6	$\|X_{in}(n)\|^5$	23	347	-35.35
7	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^2$	31	591	-37.89
8	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^3$	49	933	-37.89
9	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^4$	40	762	-37.52
10	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^5$	51	971	-37.82
11	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^2$, $\|X_{in}(n)\|^3$	17	393	-37.93
12	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^3$, $\|X_{in}(n)\|^5$	28	646	-38.25
13	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^2$, $\|X_{in}(n)\|^3$ $\|X_{in}(n)\|^4$	31	591	-38.18
14	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^2$, $\|X_{in}(n)\|^3$ $\|X_{in}(n)\|^5$	27	515	-38.12
15	$\|X_{in}(n)\|$, $\|X_{in}(n)\|^2$, $\|X_{in}(n)\|^3$ $\|X_{in}(n)\|^4$, $\|X_{in}(n)\|^5$	39	669	-38.15

where M is memory depth of the input signal, I and O are the number of neurons in the input and output layers, respectively, and H is the number of neurons in hidden layer. Here, $(H + O)$ represents the number of biases in the NN, and $((M + 1) \times I) \times H + H \times O$ represents the number of synaptic weights in the NN.

From Table II, it can be seen that the modeling performance and complexity, in terms of number of coefficients, vary with the injected envelope-dependent terms. As is well known, an NN uses a set of features for constructing a robust behavioral model [18]. The more relevant the features are that we provide to the NN as input, the greater the precision of the model [19]. This relationship is clearly illustrated in Table II. It can be seen that as the envelope term is added along with the I and Q components (second entry), performance improves relative to when only the I and Q components are provided (first entry or simply the RVFTDNN model). This improvement is achieved since the inclusion of $|X_{\text{in}}(n)|$ term is capable of generating the desired even-order intermodulation terms [20], $I(n)|X(n)|^{k-1}$ and $Q(n)|X(n)|^{k-1}$ for even k. Contrary to this effect, if only I and Q components (as in RVFTDNN) are the inputs of the NN, the network is incapable of generating the desired even-order intermodulation terms.

This outcome is described in detail in Section III-C. However, if we provide redundant terms (third to sixth entry in Table II), the performance does not improve, demonstrating the importance of the input terms provided to the NN. Entries 11–15 show that although the addition of extra terms provides reasonable performance in terms of NMSE, the complexity varies for different input terms.

In order to find the best combination/tradeoff between performance and complexity, Fig. 4 presents the complexity and NMSE curves for various injected envelope-dependent terms. From Fig. 4, it can be seen that the combination at the 11th entry of Table II offers the least complexity with relatively good performance for the nonlinear distortions of the PA. Hence, these terms were chosen for evaluating performance for the proposed method.

TABLE III
PERFORMANCE FOR POLYNOMIAL ACTIVATION FUNCTIONS

Activation function: Polynomial restricted to 3rd order (Feedforward)		
Model	Hidden Layers/Neurons per layer	NMSE (dB)
RVFTDNN	1/41	-30.04
RVFTDNN	2/13	-35.23
ARVTDNN	1/15	-37.49
Activation function: Polynomial restricted to 3rd order (Recursive)		
RVFTDNN	1/20	-34.37
ARVTDNN	1/8	-36.45

Hyperbolic tangent sigmoid functions are not easy to implement and suffer from a local minima problem as described in [20]. In order to facilitate the ease of implementation, polynomial functions restricted to second or third degree in recursion can be implemented.

For this purpose, activation function of the form $1 + x + x^2 + x^3$ was developed in MATLAB and both the RVFTDNN and the proposed models were tested using this activation function. Both feedforward and recurrent NNs (RNN) were employed to test the performance of the models. The results have been summarized in Table III. It can be seen that, for nonrecursive feedforward network, reasonable performance can be achieved by using polynomial activation functions. Similarly, using the RNN model, the desired performance can be achieved for both models. For both cases, the proposed model seems to perform better than the RVFTDNN model by around 2 dB. An important observation is that, for recursive systems, convergence can be obtained using lesser number of neurons in the hidden layer.

C. Justification of the Proposed Model

Having established that the proposed model provides better modeling capability compared to the RVFTDNN model, in terms of NMSE and number of variables, it is important to understand the factors underlying this observed improvement. The use of odd-order terms for the inverse behavioral modeling of PAs is well documented in [21], [23], and [24]. Many of the existing works do not include even-order terms (see [3]–[5], [24], [25]). The reason being that the even-order terms of the power series expansion do not lie in the first harmonic. However, as shown in [21], [23], and [24], including only odd terms is an approximation, and a better baseband model should be capable of generating both even- and odd-order terms. This combination results in a richer basis set and improved modeling accuracy and requires lower-order polynomials with better numerical properties. This finding has been validated in this paper by comparing the proposed model with the RVFTDNN.

An elaborate theoretical understanding of the importance of including the even-order terms is shown in [23]. To further clarify this matter, let us consider a baseband memoryless polynomial model described by the following expression:

$$y(n) = \sum_{j=1}^{J} a_j X(n)|X(n)|^{j-1} \tag{10}$$

where $X(n) = I(n) + jQ(n)$, $|X(n)| = (I^2(n) + Q^2(n))^{1/2}$ and J is the nonlinearity order of the polynomial. For simplicity, the memory terms in (10) have been omitted (the polynomial model shown in (25) models the memory in the system). If only odd-order terms are considered in the model, one can expand (10) to obtain

$$\begin{aligned} y_{\text{odd}}(n) &\approx (I(n) + jQ(n)) + (I(n) + jQ(n))|X(n)|^2 \\ &\quad + (I(n) + jQ(n))|X(n)|^4 + \cdots \\ &\approx (I(n) + I(n)|X(n)|^2 + I(n)|X(n)|^4 + \cdots) \\ &\quad + j(Q(n) + Q(n)|X(n)|^2 + Q(n)|X(n)|^4 + \cdots). \end{aligned} \tag{11}$$

For simplicity, the coefficients a's of the model have been ignored. When only even-order terms are considered, namely j is even, we have the following terms from (10):

$$\begin{aligned} &y_{\text{even}}(n) \\ &\approx (I(n) + jQ(n))|X(n)| + (I(n) + jQ(n))|X(n)|^3 \\ &\quad + (I(n) + jQ(n))|X(n)|^5 + \cdots \approx (I(n)|X(n)| \\ &\quad + I(n)|X(n)|^3 + I(n)|X(n)|^5 + \cdots) \\ &\quad + j(Q(n)|X(n)| + Q(n)|X(n)|^3 + Q(n)|X(n)|^5 + \cdots) \end{aligned} \tag{12}$$

where $X(n)|X(n)|^j$ represents the $(j+1)^{th}-$order nonlinearity. Here too the memory effects have also been ignored for simplicity.

As discussed earlier, a comprehensive behavioral model should be able to generate both the even- and odd-order terms. Both the RVFTDNN and the proposed ARVTDNN were investigated from this mathematical aspect. The output signal of the hidden layer's neurons was calculated. Its weights, w_{ij}^l, were set to be 1, and the memory effects of the models were ignored in the following analysis. The output of ith neuron in the first hidden layer for RVFTDNN can be written as

$$y(n) = c_i^1 = f(I(n) + Q(n) + b) \tag{13}$$

where b is the bias and f is the activation function provided in (6). This output using the hyperbolic tangent sigmoid activation function can be approximated using the Taylor series expansion shown in the following equation:

$$\begin{aligned} y(n) &= \tanh(x(n)) \\ &= x(n) - \frac{1}{3}x(n)^3 + \frac{2}{15}x(n)^5 + - \cdots \left(|x(n)| < \frac{\pi}{2}\right). \end{aligned} \tag{14}$$

For the RVFTDNN model, the input vector (ignoring memory) is $[I(n)Q(n)]$, which leads to the first- and third-order terms provided in the following equations:

$$x(n) = I(n) + Q(n) + b \tag{15}$$

where b represents the bias term added to each neuron. Using (15) and (16), as shown at the top of the next page, while considering only the important terms according to (11) and (12) and ignoring the multiplication factors, the output of the activation function of a single neuron can be approximated using (14) by

$$\begin{aligned} &y(n) \\ &= \underbrace{I(n) + Q(n)}_{\text{first order}} + \underbrace{(I(n) + Q(n))|X(n)|^2}_{\text{third order}} + \text{other terms} \\ &= (I(n) + I(n)|X(n)|^2) + (Q(n) + Q(n)|X(n)|^2) + \text{other terms}. \end{aligned} \tag{17}$$

Furthermore, the output of neurons in the hidden layer can be combined by weights, and the output signal ($I_{\text{out}}(n)$ and $Q_{\text{out}}(n)$) in the output layer can then be written as

$$I_{\text{out}}(n) = (I(n) + I(n)\,|X(n)|^2) + \text{other terms} \tag{18}$$

$$Q_{\text{out}}(n) = (Q(n) + Q(n)\,|X(n)|^2) + \text{other terms}. \tag{19}$$

It can be seen that the terms produced by the RVFTDNN model are the same for the polynomial terms when only odd-order terms are considered (11). The desired even-order terms, as shown in (12), require the square root term, i.e., $(I^2(n) + Q^2(n))^{1/2}$. It can be seen from (15)–(19) that the desired terms such as $I(n)|x(n)|$ and $Q(n)|x(n)|$, cannot be generated. These terms are important to construct the inverse AM/AM and the AM/PM characteristics of the PA.

For the ARVTDNN, the input vector, ignoring the memory terms and the square and cube of the amplitude term, is $[I(n)Q(n)|X(n)|]$. Hence, first-order term can now be written as

$$x(n) = I(n) + Q(n) + |X(n)| + b \tag{20}$$

and the third-order term is presented at the top of the next page (21). Similar to (17), using (20) and (21), the output of

$$\begin{aligned} x^3 &= (I(n)+Q(n)+b)^3 \\ &= I^3(n)+Q^3(n)+3I(n)Q^2(n)+3I^2(n)Q(n)+3b(I^2(n)+Q^2(n)+2I(n)Q(n))+3b^2(I(n)+Q(n))+b^3 \\ &= I(n)\,|X(n)|^2+Q(n))\,|X(n)|^2+3b(|X(n)|^2+2I(n)Q(n))+3b^2(I(n)+Q(n))+\text{other terms} \end{aligned} \tag{16}$$

$$\begin{aligned} x^3 &= (I(n)+Q(n)+|X(n)|+b)^3 \\ &= (I(n)+Q(n)+|X(n)|)^3+3b(I(n)+Q(n)+|X(n)|)^2+3b^2(I(n)+Q(n)+|X(n)|)+b^3 \\ &= 4(I(n)+Q(n))\,|X(n)|^2+6b(I(n)+Q(n))\,|X(n)|+3b^2(I(n)+Q(n)+|X(n)|)+\text{other terms} \end{aligned} \tag{21}$$

the activation function using (14) can then be approximated by

$$\begin{aligned} &y(n) \\ &= \underbrace{I(n)+Q(n)}_{\text{first order}}+\underbrace{(I(n)+Q(n))\,|X(n)|}_{\text{second order}} \\ &\quad+\underbrace{(I(n)+Q(n))\,|X(n)|^2}_{\text{third order}}+\text{other terms} \\ &= (I(n)+I(n)\,|X(n)|+I(n)\,|X(n)|^2) \\ &\quad+(Q(n)+Q(n)\,|X(n)|+Q(n)\,|X(n)|^2)+\text{other terms}. \end{aligned} \tag{22}$$

Similarly, adjusting the weights and biases leads to the following output components:

$$\begin{aligned} &I_{\text{out}}(n) \\ &\quad= I(n)+I(n)\,|X(n)|+I(n)\,|X(n)|^2+\text{other terms} \end{aligned} \tag{23}$$

$$\begin{aligned} &Q_{\text{out}}(n) \\ &\quad= Q(n)+Q(n)\,|X(n)|+Q(n)\,|X(n)|^2+\text{other terms}. \end{aligned} \tag{24}$$

The NN's output signal can be written as $y_{\text{out}}(n) = I_{\text{out}}(n)+jQ_{\text{out}}(n)$; hence, $(I(n)+jQ(n))|X(n)|$ and $(I(n)+jQ(n))|X(n)|^2$ are the second- and third-order terms of the input signal, respectively. Comparing (23) and (24) with (11) and (12) shows that the proposed model is able to generate a richer basis function set that includes the desired even-order nonlinear terms: $I(n)|x(n)|$ and $Q(n)|x(n)|$ needed for the synthesis of accurate and effective DPD model without the need to make the NN recursive to obtain a better approximation of both odd and even terms. This avoids the issue of over fitting and instability that might arise form a closed-loop system.

The improved performance for the proposed model, in terms of NMSE and adjacent channel power ratio (ACPR), as will be shown in Section V, can be explained using this analysis. By providing better and relevant features to the NN, it can generate a richer basis function set without the need for recursion and results in reducing the modeling complexity of the PA [18], which can be seen from the comparison between the second and seventh entries of Table II with the 11th entry. When these envelope-dependent terms, ($|X(n)|^2$, $|X(n)|^3$) are injected into the DPD, better cancelation of the nonlinear distortions can be achieved as higher order even- and odd-order products are generated without adding additional hidden layers in the network.

IV. Measurement Setup and Transmitter Performance

A. Devices/Signals in Measurement Setup

The PA in the measurement setup is a highly nonlinear Doherty PA with a SSG of 13 dB and a saturation power (P_{sat}) of 44 dBm. The PA was driven by a modulated wideband long-term evolution-advanced (LTE-advanced) signal with a bandwidth of 60 MHz and a PAPR of 10.4 dB. The signal was modulated at a center frequency of 2.14 GHz. The corresponding sampling frequency of the signal was 384 MHz. The duration time of the signal was 1 ms.

B. Measurement Setup and Model Identification

Fig. 5(a) and (b), respectively, presents the measurement setup and the identification procedure that was used for the acquisition of the Doherty PA's output signal and the inverse modeling procedure. The input signal was first synthesized by using advanced design system (ADS) software and then downloaded into the arbitrary waveform generator (AWG 81180A) through an Internet port. The baseband signal generated by the AWG was transmitted via a cable to the performance signal generator (PSG E8267D) which performs digital modulation, digital-to-analog conversion, and frequency up-conversion.

The generated RF signal was then fed to the driver. The signal power at the output of the PA was decreased by the attenuator before being captured by the high sampling rate oscilloscope (MSO 9404A). The entered RF signal in the MSO 9404A was analyzed using Keysight 89600 vector signal analyzer (VSA) software, which runs on the MSO 9404A. The VSA also performs frequency down-conversion, analog-to-digital conversion, and demodulation. The digital output baseband signal was then obtained using the VSA. The acquired output and input signals were processed in MathWork's MATLAB software, in order to construct the behavioral model of the transmitter.

The DPD using the proposed algorithm was carried out using the indirect learning algorithm (ILA) [26]. Fig. 5(b) shows the procedure for the inverse modeling of the nonlinear characteristics of the transmitter. In the beginning, the DPD block was not used, and both the input and the output of the transmitter were required to generate the reference model (post inverse). The error was computed between the modeled output and the desired output. Hence, the time alignment between the input and output signals needed to be done first, which could be achieved using a cross correlation technique [27].

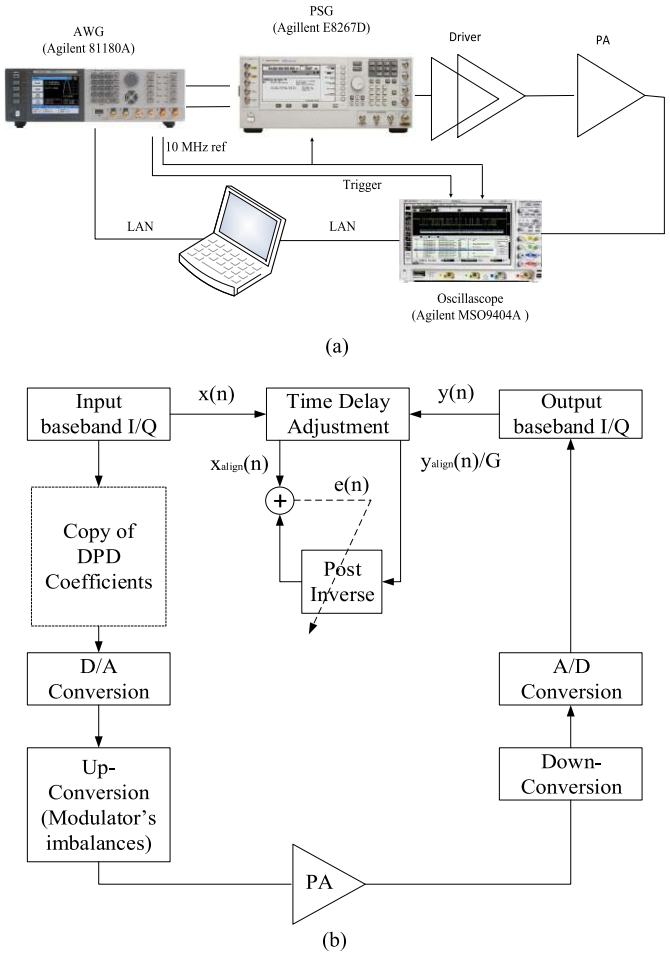

Fig. 5. (a) Block diagram of a measurement setup. (b) Flow diagram of the modeling procedure using ILA.

In order to more accurately construct the behavioral model of the transmitter, the output signal is usually normalized by the SSG [G in [Fig. 5(b)] of the Doherty PA, because the normalized output signal can improve modeling performance. Once the error has converged to minimum, the coefficients obtained using the post inverse was placed in the DPD block to obtain the desired linearization.

C. Different Cases for Performance Evaluation

In this paper, we present three cases of the direct-conversion transmitter as shown in Table IV. Case 1 was a condition in which the nonlinear distortions of the PA only existed in the transmission chain. Case 2 had the presence of both the I/Q imbalance and PA nonlinear distortions. For Case 3, the dc offset, I/Q imbalance, and PA nonlinear distortions were all present in the transmitter, with dc offset values of 3% and 5% in the I and Q branches, respectively. The amplitude imbalance was 1 dB, and the I/Q phase imbalance was 5°.

The gain and phase characteristics of the transmitter for different cases are presented in Fig. 6(a) and (b), respectively. From Fig. 6, it can be seen that Case 1 not only had nonlinear characteristics at the high power level, but also had strong memory effects that resulted in massive dispersion.

TABLE IV
DIFFERENT CONDITIONS OF THE TRANSMITTER SYSTEM'S IMPAIRMENTS

	Various Distortions	Distortion Level
Case 1	PA nonlinearity: ✓ *I*/*Q* imbalance: × DC offsets: ×	PA nonlinearity: 3 dB gain compression
Case 2	PA nonlinearity: ✓ *I*/*Q* imbalance: ✓ DC offsets: ×	PA nonlinearity: Same as Case1 *I*/*Q* imbalance: 1 dB gain compression and 3 dB phase compression
Case 3	PA nonlinearity: ✓ *I*/*Q* imbalance: ✓ DC offsets: ✓	PA nonlinearity: Same as Case1 *I*/*Q* imbalance: Same as Case 1 DC offsets: 3% for *I* and 5% for *Q*

These memory effects were due to the wide bandwidth and energy-storing elements in the bias circuit of the amplifier [7].

The gain and phase characteristics of Cases 2 and 3 are also presented in Fig. 6 to illustrate the effects of the I/Q imbalance and dc offset, which tremendously altered the gain and phase characteristics of the transmitter. This situation led to a corresponding change in DPD model, because the DPD model is the inverse model of the transmitter. If the DPD model cannot follow the change of the transmitter, the linearization performance of the DPD model deteriorates dramatically.

V. MEASUREMENT RESULTS

In order to evaluate the performance of the proposed DPD model, the DPD model of the transmitter was built using the proposed ARVTDNN and other methods presented in the literature, including the RVFTDNN model, the conjugate memory polynomial model (CMPM) [28], [29], and the memory polynomial model (MPM) [4].

The same algorithm and parameters, such as the number of hidden layers and transfer function, were used to construct the behavioral model of the transmitter for the RVFTDNN [8] as with the ARVTDNN. The RVFTDNN had only I and Q components as shown in the first entry of Table II. The optimal number of neurons in the hidden layer for RVFTDNN was obtained using the same method as ARVTDNN.

The MPM is a popular method for the modeling and mitigation of nonlinear distortions in PAs. The MPM can be expressed as [4]

$$y_{\mathrm{MPM}}(n) = \sum_{j=1}^{K} \sum_{i=0}^{M} a_{\mathrm{ij}} x(n-i)\,|x(n-i)|^{j-1} \qquad (25)$$

where $x(n)$ is the input signal of a DUT, a's are the coefficients of the model, $y(n)$ is the output signal of a DUT, and M and K represent the memory depth and nonlinearity order, respectively.

The CMPM is very useful for the compensation of various impairments in the direct-conversion transmitter such as PA nonlinearity and modulator's imperfections. The CMPM with

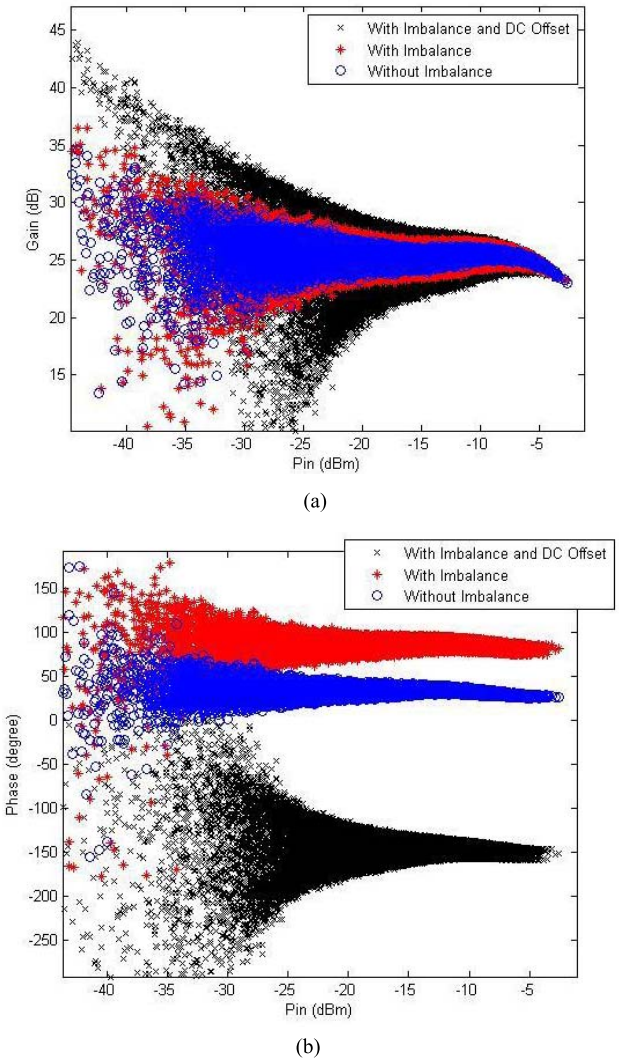

Fig. 6. Characteristics of the transmitter system. (a) Gain characteristics. (b) Phase characteristics.

nonlinearity orders K_1 and K_2 and memory depth M_1 can be represented as [28]

$$y_{\text{CMPM}}(n) = \sum_{j_1=1}^{K_1}\sum_{i_1=0}^{M_1} a_{i_1 j_1} x(n-i_1)|x(n-i_1)|^{j_1-1} + \sum_{j_2=1}^{K_2}\sum_{i_2=0}^{M_2} a_{i_2 j_2} x^*(n-i_2)|x^*(n-i_2)|^{j_2-1} + e \quad (26)$$

where e is a constant for dc offset compensation and $x^*(n)$ is the complex conjugate of the $x(n)$.

A. NMSE Performance, Numerical Stability, and Complexity

Once the experimental setup was complete and various cases of evaluation were developed, the next steps were the evaluation of the proposed methodology for the aforementioned cases and an extensive comparison with existing state-of-the-art methods.

TABLE V
NMSE PERFORMANCE OF VARIOUS MODELS WITH THE OPTIMAL VARIABLE PARAMETERS FOR DIFFERENT CASES

	Case 1 (dB)	Case 2 (dB)	Case 3 (dB)
MPM	-34.83	-19.56	-16.38
CMPM	-37.38	-37.12	-36.84
ARVTDNN	-37.93	-37.61	-37.53
RVFTDNN	-35.72	-35.63	-35.26

An important figure of merit to evaluate the in-band performance of the proposed methodology is the NMSE given by (8). Table V presents the NMSE performances of ARVTDNN, RVFTDNN, CMPM, and MPM in the context of optimal parameters for Cases 1, 2, and 3.

From Table V, it can be seen that all the four models had good modeling capability for Case 1, i.e., when only the PA's nonlinear distortions were considered, and the modulator's impairments were ignored. However, for Cases 2 and 3, the MPM was incapable of mitigating the joint impairments of the PA and the modulator. This poor performance was due to the inability of the MPM, in its present form, to compensate for the direct conversions transmitter's imperfections other than PA nonlinearity.

The CMPM was, therefore, introduced, where a conjugate term along with a dc term were appended to the conventional MPM model. It can be seen from Table V that the CMPM and RVFTDNN were capable of modeling the gain and phase imbalances along with the dc offset and PA's nonlinearity, resulting in a reduced NMSE value.

The proposed ARVTDNN model was also capable of mitigating these imperfections. In addition, it can be seen that the ARVTDNN model exhibited the best modeling performance among four models for all cases.

Bias or "systematic error" can be defined as the deviation between the estimated parameters and the actual parameters. In this paper, the NMSE is a reasonable measure of the bias present in the system. Fig. 7 illustrates the spectrum of the residual error (in dBm/Hz) after modulator impairment mitigation and PA linearization. There are various sources of residual error, namely, modeling error and measurement noise. The reason for the modeling error is that the predistorter aims to model the exact inverse of the nonlinear system. However, the model provides an approximation and the true inverse may not be reached, resulting in a modeling error.

In addition, other impairments in the transceiver lead to measurement noise. These impairments are generated from the imperfect behavior of other components present in the transceiver, such as the analog-to-digital and digital-to-analog

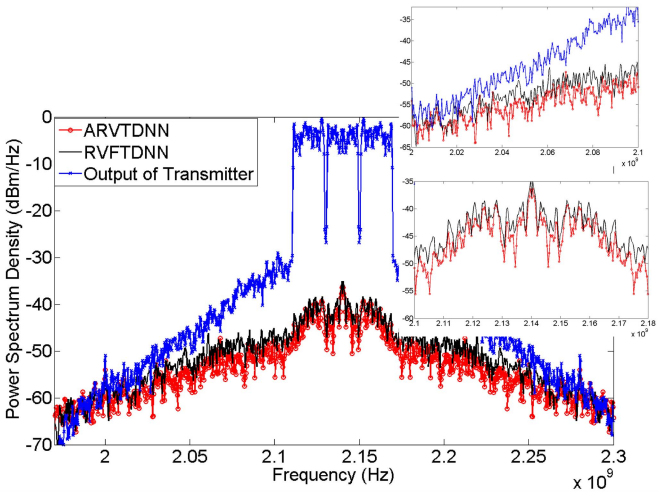

Fig. 7. Spectrum error analysis (PSD of the modeling error) for the proposed model and RVFTDNN model for Case 3.

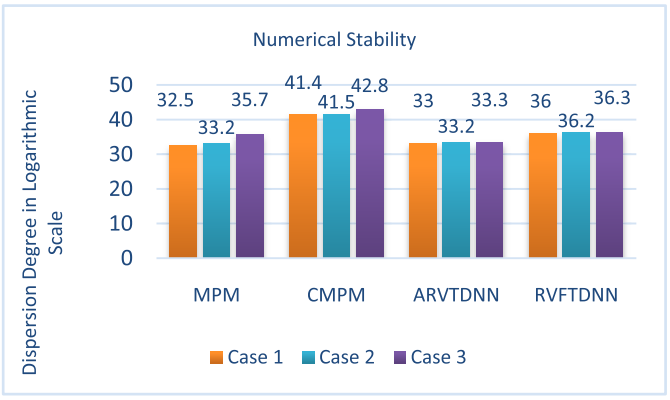

Fig. 8. Evaluation of numerical stability and dispersion degree.

converters, losses in the feedback path, and imperfections in the signal generator and spectrum analyzer. These impairments lead to errors in the measurements. In our case, since the channel between the transmitter and receiver was considered to be ideal, the main source of the noise was the measurement noise.

It is shown in Fig. 7 that the in-band and out-of-band errors for the proposed model were lower than those of the RVFTDNN. These values indicate the efficacy of the proposed methodology.

Numerical stability is also an important feature for evaluation of the performance of the transmitter's inverse modeling algorithm [30]. A key measure of numerical stability is the dispersion degree of the modeling coefficients, which is defined as [30]

$$\alpha = \frac{\max(|a_{ij}|)}{\min(|a_{ij}|)} \tag{27}$$

where a_{ij} represents the modeling coefficients.

Fig. 8 shows the dispersion degree of the modeling coefficients for different modeling algorithms and cases. From Fig. 8, it can be seen that the CMPM, despite its good modeling capability, had the poorest numerical stability of all the models in terms of the dispersion degree of its coefficients. Moreover, CMPM required matrix inversion; and, due to the appending of the conjugate terms, the size of the matrix increased significantly. As a result, the matrix conditioning deteriorated, and its inversion becomes highly unstable.

In comparison with polynomial-based models (MPM and CMPM), the NN-based models exhibited better dispersion coefficients. In addition, the ARVTDNN had a 3 dB (twice in linear scale) improved performance in terms of dispersion coefficient as compared to its counterpart, the RVFTDNN model.

Another important measure in the evaluation of the performance of the modeling algorithm is running complexity, as measured by the number of parameters and floating-point operations (FLOPs) [31]. The numbers of parameters for the four models are shown in Table VI. For the MPM model, the number of variables can be calculated as $(M+1)\times K$. For the CMPM model, the number of variables are $(M_1+1)\times K_1+(M_1+1)\times K_2+1$. Similarly, (9) provides the expression for determining the number of variables for the RVFTDNN and ARVFTDNN models.

From Table VI, it can be seen that the ARVTDNN had less complexity than the RVFTDNN, while achieving better performance. The percentage reduction in complexity can be calculated as $(453-393)/453\times 100\% = 13\%$. For polynomial methods, the MPM had less complexity than the CMPM; however, the CMPM had better NMSE performance than MPM as shown in Table V.

The running complexities (based on FLOPs) of the ARVTDNN and RVFTDNN models in Case 2 are shown in Fig. 9. Currently, there are three basic components—addition, multiplication, and activation functions—for constructing the inverse model of the PA using NNs. Hence, these three components were used for comparisons of running complexity. This technique has been elaborated in [31]. The FLOPs of the three basic components were based on the number of nodes in different layers. The numbers of FLOPs considering both addition and multiplication were calculated by $((M+1)\times I)\times H + H\times O$, and the number of activation function was $H+O$. From Fig. 8, it can be seen that the ARVTDNN had a lower number of additions and

TABLE VI

PERFORMANCE OF VARIOUS MODELS IN TERMS OF NMSE, OPTIMAL PARAMETERS, AND TOTAL NUMBER OF VARIABLES

	Optimal parameters	Total no. of variables
MPM	$M=7, K=12$	96
CMPM	$M_1=7, K_1=12, K_2=12$	193
ARVTDNN	$M=3, I=5, H=17, O=2$	393
RVFTDNN	$M=3, I=2, H=41, O=2$	453

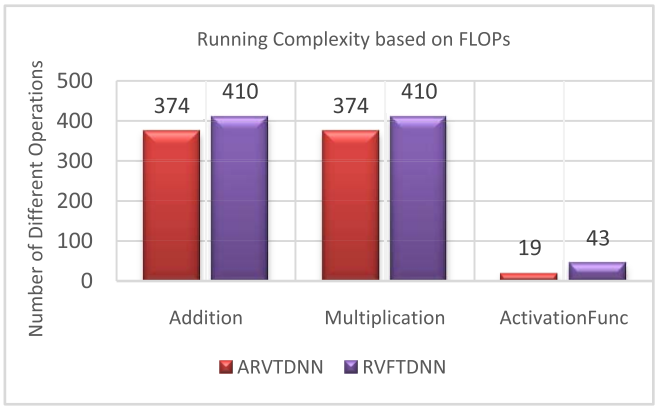

Fig. 9. Running complexity of the ARVTDNN and RVFTDNN models based on FLOPs.

TABLE VII
ACPR PERFORMANCE OF VARIOUS MODELS FOR DIFFERENT CASES

	Case 2 (Lower/upper)	Case 3 (Lower/upper)
NO-DPD (dB)	-31.55/-31.17	-31.68/-31.47
MPM (dB)	-41.63/-39.98	-41.99/-40.18
CMPM (dB)	-44.95/-44.74	-44.39/-44.54
RVFTDNN (dB)	-44.48/-43.67	-43.85/-44.34
ARVTDNN (dB)	-47.51/-47.65	-47.26/-47.19

multiplications than the RVFTDNN. The number of additions and multiplications were reduced by 8.8% while the number of activation functions was reduced by 55.8%.

B. Linearization Performance for Various Models

Table VII displays the ACPR performances of various models for Cases 2 and 3. The ACPR was evaluated in this paper with a carrier frequency and frequency offset of 2.14 GHz and 40 MHz, respectively, and the evaluation bandwidth of 18 MHz. From Table VII, it can be seen that in the transmitter linearized by the ARVTDNN model, the ACPR was better than the no-DPD model by 16 dB and the RVFTDNN model by 3–4 dB. Thus, the ARVTDNN model was superior to all other models for the compensation of impairments in the transmitter.

Figs. 10 and 11 show the power spectral density (PSD) of the transmitter linearized by the ARVTDNN, RVFTDNN, CMPM, and MPM for Cases 2 and 3. From Figs. 10 and 11, it can be clearly seen that the imperfect transmitter not only generated very high out-of-band distortions in Cases 2 and 3, but also engendered a dc component at the carrier frequency for Case 3.

The MPM did not have good compensation capability in the presence of the I/Q imbalance and dc offset. However, the CMPM provided good compensation for the I/Q imbalance and dc offset, which were compensated by the appended complex conjugate term and constant term shown in (26). The CMPM showed better modeling capability than the RVFTDNN model. However, in comparison to other models, the proposed ARVFTDNN model had the best linearization capability.

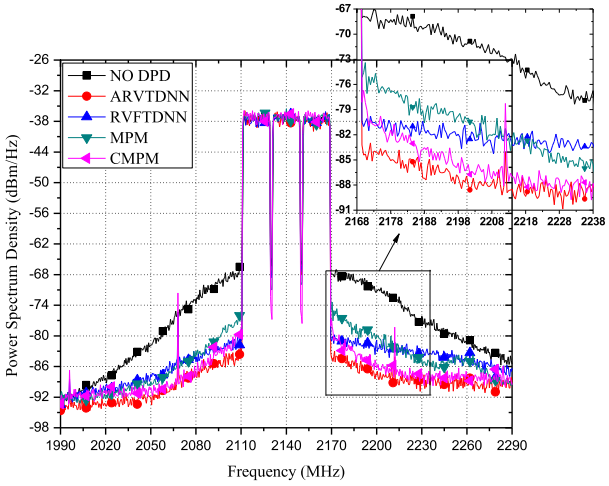

Fig. 10. Linearization performance of the transmitter using various models for Case 2.

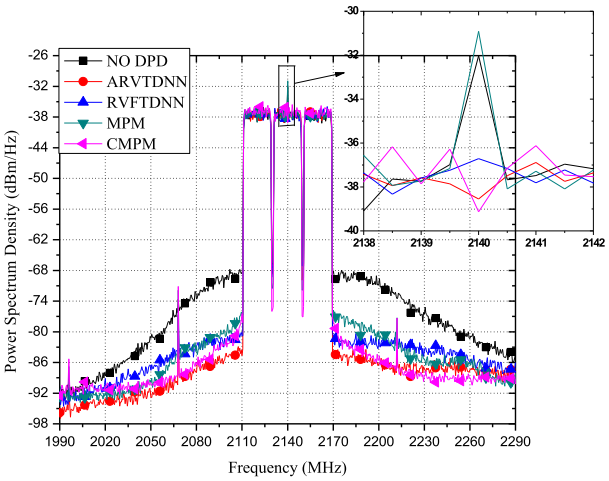

Fig. 11. Linearization performance of various models for the PA model, dc offset, and I/Q imbalance.

VI. CONCLUSION

This paper proposes a digital predistorter modeled by an ARVTDNN that can be used for mitigation of the distortions generated by a direct-conversion transmitter. The ARVTDNN is a one-step solution for the dc offset, I/Q imbalance, and nonlinear PA distortions, and is capable of generating both even- and odd-order intermodulation products, resulting in improved performance.

A comparative study of ARVTDNN and other existing models was carried out for the compensation of transmitter imperfections. The simulation results show that the linearization performance of the ARVTDNN, in terms of NMSE, was better than the RVFTDNN model by 2–3 dB. The measurement results reveal that the ARVTDNN's linearization capability, in terms of ACPR, was better than that of RVFTDNN by 3–4 dB. At the same time, the ARVTDNN had less running complexity, based on the number of parameters and FLOPs, as well as having stronger numerical stability (equal to 3 dB) than the RVFTDNN. The superior linearization performance

of the ARVTDNN in contrast to the other polynomial models considered was demonstrated.

As mentioned earlier, the DPD in the proposed mitigation strategy uses indirect learning architecture (ILA) [26]. In the future, direct learning architecture (DLA) [26] can be used to alleviate the aforementioned transmitter's imperfections. Using the current ILA-based strategy, the DPD coefficients once placed will not be able to adapt to the changes in the operating condition of the system unless updated using new batch-data-based iteration of the model identification procedure. A DLA will be able, in principle, to adapt better and faster to the changes in the system's conditions, such as the input signal's average power, heating effects, and aging.

REFERENCES

[1] J. Kim, Y. Y. Woo, J. Moon, and B. Kim, "A new wideband adaptive digital predistortion technique employing feedback linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 2, pp. 385–392, Feb. 2008.

[2] A. Ghadam, S. Burglechner, A. H. Gokceoglu, M. Valkama, and A. Springer, "Implementation and performance of DSP-oriented feed-forward power amplifier linearizer," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 2, pp. 409–425, Feb. 2012.

[3] H. H. Boo, S. Chung, and J. L. Dawson, "Digitally assisted feed-forward compensation of Cartesian-feedback power-amplifier systems," *IEEE Trans. Circuits Syst. II, Exp. Brief*, vol. 58, no. 8, pp. 457–461, Aug. 2011.

[4] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.

[5] O. Hammi, A. Kwan, S. Bensmida, K. A. Morris, and F. M. Ghannouchi, "A digital predistortion system with extended correction bandwidth with application to LTE-A nonlinear power amplifiers," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 12, pp. 3487–3495, Dec. 2014.

[6] L. Ding, Z. Ma, D. R. Morgan, M. Zierdt, and G. T. Zhou, "Compensation of frequency-dependent gain/phase imbalance in predistortion linearization systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 55, no. 1, pp. 390–397, Feb. 2008.

[7] F. Mkadem, "Behavioural modeling and linearization of RF power amplifier using artificial neural networks," M.S. thesis, Dept. Elect. Eng., Univ. Waterloo, Waterloo, ON, Canada, 2010.

[8] M. Rawat and F. M. Ghannouchi, "A mutual distortion and impairment compensator for wideband direct-conversion transmitters using neural networks," *IEEE Trans. Broadcast.*, vol. 58, no. 2, pp. 168–177, Jun. 2012.

[9] S. Lajnef, N. Boulejfen, A. Abdelhafiz, and F. M. Ghannouchi, "Two-dimensional Cartesian memory polynomial model for nonlinearity and I/Q imperfection compensation in concurrent dual-band transmitters," *IEEE Trans. Circuits Syst. II, Exp. Brief*, vol. 63, no. 1, pp. 14–18, Jan. 2016.

[10] J. K. Cavers, "New methods for adaptation of quadrature modulators and demodulators in amplifier linearization circuits," *IEEE Trans. Veh. Technol.*, vol. 46, no. 3, pp. 707–716, Aug. 1997.

[11] Y.-D. Kim, E.-R. Jeong, and Y. H. Lee, "Adaptive compensation for power amplifier nonlinearity in the presence of quadrature modulation/demodulation errors," *IEEE Trans. Signal Process*, vol. 55, no. 9, pp. 4717–4721, Sep. 2007.

[12] C. L. P. Chen, J. Wang, C.-H. Wang, and L. Chen, "A new learning algorithm for a fully connected neuro-fuzzy inference system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1741–1757, Oct. 2014.

[13] 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.

[14] M. Robnik-Šikonja, "Data generators for learning systems based on RBF networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 5, pp. 926–938, May 2016.

[15] M. Ibukahla, J. Sombria, F. Castanie, and N. J. Bershad, "Neural networks for modeling nonlinear memoryless communication channels," *IEEE Trans. Commun.*, vol. 45, no. 7, pp. 768–771, Jul. 1997.

[16] N. Benvenuto, F. Piazza, and A. Uncini, "A neural network approach to data predistortion with memory in digital radio systems," in *Proc. IEEE Int. Conf. Commun.*, Singapore, vol. 1, Nov. 1995, pp. 152–156.

[17] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.

[18] Y. M. Mohmad Hassim and R. Ghazali, "Training a functional link neural network using an artificial bee colony for solving a classification problems," *J. Comput. Press*, vol. 4, pp. 110–115, Dec. 2012.

[19] S. Ledesma, G. Cerda, G. Aviña, D. Hernández, and M. Torres, "Feature selection using artificial neural networks," in *Advances in Artificial Intelligence*, A. Gelbukh, Ed. Mexico, North America: Springer, 2008.

[20] R. Rojas, *Neural Networks: A Systematic Introduction*. Berlin, Germany: Springer, 1996.

[21] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Augmented Hammerstein predistorter for linearization of broad-band wireless transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 4, pp. 1340–1349, Jun. 2006.

[22] M. Hui, T. Liu, M. Zhang, Y. Ye, D. Shen, and X. Ying, "Augmented radial basis function neural network predistorter for linearisation of wideband power amplifiers," *Electron. Lett.*, vol. 50, no. 12, pp. 877–879, Jun. 2014.

[23] H. Enzinger, K. Freiberger, and C. Vogel, "Analysis of even-order terms in memoryless and quasi-memoryless polynomial baseband models," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 1714–1717.

[24] L. Ding and G. T. Zhou, "Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 156–162, Jan. 2004.

[25] E. Westesson and L. Sundstrom, "A complex polynomial predistorter chip in CMOS for baseband or IF linearization of RF power amplifiers," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 1999, pp. 206–209.

[26] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *Proc. IEEE Int. Symp. Wireless Commun. Syst.*, Oct. 2008, pp. 309–313.

[27] M. S. O. Alink, E. A. M. Klumperink, M. C. M. Soer, A. B. J. Kokkeler, G. J. M. Smit, and B. Nauta, "A CMOS-compatible spectrum analyzer for cognitive radio exploiting crosscorrelation to improve linearity and noise performance," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 3, pp. 479–492, Mar. 2012.

[28] F. M. Ghannouchi, O. Hammi, and M. Helaoui, *Behavioral Modeling and Predistortion of Wideband Wireless Transmitters*. Hoboken, NJ, USA: Wiley, 2015.

[29] L. Anttila, P. Handel, and M. Valkama, "Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 730–739, Apr. 2010.

[30] M. Rawat, F. M. Ghannouchi, and K. Rawat, "Three-layered biased memory polynomial for dynamic modeling and predistortion of transmitters with memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 60, no. 3, pp. 768–777, Mar. 2013.

[31] L. Mingyu, J. Liu, Y. Jiang, and W. Feng, "Complex-Chebyshev functional link neural network behavioral model for broadband wireless power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1979–1989, Jun. 2012.

Dongming Wang received the B.Sc. degree in electrical engineering from the Nanyang Institute of Technology, Nanyang, China, in 2013, and the master's degree from the Department of Electrical and Computer Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada, in2016.

He was with the iRadio Laboratory, Department of Electrical and Computer Engineering, University of Calgary. His current research interests include the application of neural networks for digital predistortion systems.

Mohsin Aziz received the M.Sc. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 2013, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Schulich School of Engineering.

He is associated with the iRadio Laboratory, Department of Electrical and Computer Engineering, University of Calgary. His current research interests include data-aided and blind digital signal processing techniques for modeling and mitigation of imperfections in transceivers.

Mohamed Helaoui (S'06–M'09) received the M.Sc. and Ph.D. degrees in communications and information technology from the Ecole Supérieure des Communications de Tunis, Ariana, Tunisia, in 2003 and 2006, respectively, and the Ph.D. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 2008.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of Calgary. He has authored or co-authored over150 publications, one book, and three book chapters. He holds 11 patents applications (seven of them are granted). His current research interests include digital signal processing, power efficiency enhancement for wireless transmitters, efficient and broadband power amplifiers (PAs), monolithic microwave-integrated circuit PAs for wireless and satellite communications, six-port receivers and advanced transceiver design for software defined radio, and millimeter wave applications.

Dr. Helaoui is the Chair of the IEEE local COM/MTT/AP Joint Chapter in the Southern Alberta Section.

Fadhel M. Ghannouchi (F'07) joined the Ecole Polytechnique de Montreal, Montreal, QC, Canada, in 2005, where he has been teaching microwave theory and techniques and RF communications systems since 1984. He has held several invited positions at several academic and research institutions in Europe, North America, and Japan. He has provided consulting services to a number of microwave and wireless communications companies. He is currently a Professor, iCORE/Canada Research Chair, and the Director of the iRadio Laboratory, Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB, Canada. He has authored over 500 refereed publications, holds seven U.S. patents, and seven patents applications. His current research interests include RF and wireless communications, nonlinear modeling of microwave devices and communications systems, design of power and spectrum efficient microwave amplification systems, and design of SDR systems for wireless and satellite communications applications.

Dr. Ghannouchi is an IET Fellow and the IEEE-MTT-S Distinguished Microwaves Lecturer.