IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2019, accepted July 6, 2019, date of publication July 10, 2019, date of current version July 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927875

Vector Decomposition Based Time-Delay Neural Network Behavioral Model for Digital Predistortion of RF Power Amplifiers

YIKANG ZHANG[1,2], YUE LI[3], (Student Member, IEEE), FALIN LIU[1,2], AND ANDING ZHU[3], (Senior Member, IEEE)

[1]Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China
[2]Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei 230027, China
[3]School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, D04 V1W8 Ireland

Corresponding author: Falin Liu (liufl@ustc.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant Number 61471333, and in part by the Science Foundation Ireland under Grant 13/RC/2077 and Grant 17/NSFC/4850.

ABSTRACT This paper presents two novel neural network models for radio-frequency (RF) power amplifiers (PAs): vector decomposed time-delay neural network (VDTDNN) model and augmented vector decomposed time-delay neural network (AVDTDNN) model. In contrast to conventional neural network-based models, VDTDNN and AVDTDNN comply with the physical characteristics of RF PAs by employing carefully designed network structures. In particular, the nonlinear operations are conducted only on the magnitude of the input signals, while the phase information is recovered with the linear weighting. Linear terms with shortcut connection, as well as high-order terms, can be used to further boost the modeling performance. The complexity analysis shows that the proposed models have significantly lower complexity than the existing neural network models. A wideband GaN RF PA excited by the 40- and 60-MHz OFDM signals were employed to evaluate the performance. The extensive experimental results reveal that the proposed VDTDNN and AVDTDNN models can achieve better linearization performance with lower computational complexity compared with the existing neural network-based models.

INDEX TERMS Nonlinear RF PA, digital predistortion, artificial neural network, vector decomposition, behavioral modeling.

I. INTRODUCTION

Driven by the consumer demand and wireless communication technology evolution, the transmitted signals in modern communication systems have the tendencies towards higher peak-to-average power ratios (PAPR) and wider bandwidth [1], [2], which deteriorates the bit error rate (BER) and intensifys the adjacent channel interference effects, due to the inherent nonlinear characteristics of radio frequency (RF) power amplifier (PA). In order to resolve the conflict between linearity and efficiency of RF PA, many linearization approaches including feed-forward, analog predistortion and digital predistortion (DPD) were developed [3], wherein DPD is widely considered as the most promising technology due to its high flexibility and relatively low power consumption. The core idea of DPD is to extract an inverse behavioral model in digital domain for the nonlinear RF PA, then cascade the predistorter in the forward baseband. Consequently, the cascade of the nonlinear inverse model and RF PA will be a linear system [4]. By properly selecting the models for digital predistorter in the forward path, the nonlinearity of RF PA can be effectively compensated [5]–[7].

Many DPD models have been developed to address the nonlinearity of RF PA [3], [8]–[10], where Volterra-based models are the most widely used models. However, the drawback of the Volterra-based model is the high correlation between different polynomial basis functions, which limits the performance improvement even when the number of polynomial terms increase dramatically [11]. In addition, for the sake of meeting industry requirements, many advanced RF PA architectures, such as out-phasing PA, distributed PA and multistage Doherty PA, have been proposed, and some of

The associate editor coordinating the review of this manuscript and approving it for publication was Vittorio Camarchia.

their behavioral characteristics may be very different with the traditional characteristics [12], so more flexible models should be developed to satisfy the rapidly developing industrial demands.

Due to its excellent capability of nonlinear fitting [13], neural network (NN) has been considered as a promising method for DPD and many DPD models based on NN have been developed so far. Two complex-valued multilayer perceptron (MLP) was employed to model the AM-AM and AM-PM characteristics separately in [14], two drawbacks of this method are as follows: (i) the nonlinear mapping of RF PA was divided into two independent networks, and it is difficult for them to converge to the optimal states simultaneously. (ii) as the complex-valued coefficients should be updated with complex gradient operations, the computational complexity is too high [15]. In order to avoid the complex gradient operations, a real-valued time-delay neural network (RVTDNN) model was reported in [16], where both input and output were split into in-phase and quadrature (IQ) parts to implement the nonlinear model. After that, with the same input and output strategy, many neural network models have been developed to implement DPD [17]–[20]. A composite neural network DPD model for multi-input multi-output (MIMO) transmitter was developed in [21], where the crosstalk effects, PA nonlinearity and IQ imbalance were compensated simultaneously. Recently, the performance of deep neural network with different activation functions was investigated in [11]. Not only PA's nonlinearity and IQ imbalance, but also frequency, temperature and voltage variations were considered in DPD, the neural network DPD was implemented for mobile radio communications in [22], where the experimental results show that the performance can meet the telecommunication standards.

Although different neural network topologies have been developed to implement DPD, almost all of the existing neural network DPD models split the input and output signals into I and Q parts to model the nonlinear behavioral characteristics of RF PAs, which can violate the "first-zone constraint" [23], [24] and thus leads to limited performance. Instead, in this work, we propose two new neural network models: vector decomposed time-delay neural network (VDTDNN) model and augmented vector decomposed time-delay neural network (AVDTDNN) model and they conform more with the nonlinear physical mechanisms of RF PAs, where only the envelops of the input signal are conducted nonlinear operations while the phase information is recovered with linear weighting operations. Both theoretical analysis and experimental results reveal that the performance and computational complexity of the proposed models are superior to the existing neural network DPD models.

The rest of the paper is organized as follows. Section II describes the proposed VDTDNN model. The Proposed AVDTDNN is introduced in Section III. A computational complexity comparison is analyzed in Section IV. The experimental validation and a brief conclusion are presented in Section V and Section VI, respectively.

II. VECTOR DECOMPOSED TIME-DELAY NEURAL NETWORK

In view of physical mechanisms of nonlinear operations, the complex signals (I+jQ) only exist in baseband and only real valued signals would be sent into RF PA after up-conversion [25], it means that the nonlinear operations are operated on envelope signals (including both I and Q components). Accordingly, the designed RF PA model should conduct nonlinear operations on envelope signals rather than dividing signals into real and imaginary parts independently to conduct nonlinear operations. Besides, two other constraints should be satisfied: (i) the designed models should meet the "first-zone constraint", where odd-parity and unitary phase constraints must be satisfied [23], [24]. (ii) the models can handle complex-valued signals avoiding complex gradient operations.

A. DESIGN VECTOR DECOMPOSED SUB-NETWORK

Generally, in RF PAs, nonlinearities are mainly induced by AM-AM and AM-PM distortion. To build a model for the RF PA, the math expression can be expressed as (1), where nonlinear operations are conducted by using magnitude and then the phase information is recovered with linear weighting of $e^{j\boldsymbol{\theta}_g}$:

$$\tilde{y}(n) = \sum_{g=1}^{G} F_g[\cdot] = \sum_{g=1}^{G} \tilde{a}_g \cdot A_g \cdot e^{j\boldsymbol{\theta}_g} \tag{1}$$

where $\tilde{y}(n)$ is the complex based output signal at instant n. $F_g[\cdot] = \tilde{a}_g \cdot A_g \cdot e^{j\boldsymbol{\theta}_g}$ is the gth basis function, where $\tilde{a}_g = a_{gI} + ja_{gQ}$ is the complex coefficient, a_{gI} and a_{gQ} are the real and imaginary parts of the coefficient. A_g represents the nonlinear operation output of the gth basis function for input sequence $(\tilde{x}(n),\tilde{x}(n-1),\cdots,\tilde{x}(n-M))$. M is the memory depth and G is the number of basis function. $e^{j\boldsymbol{\theta}_g}$ includes the recovered phase information for the gth basis function.

If we only consider one of the basis functions $F_g[\cdot]$ in (1), it yields:

$$\begin{aligned} F_g[\cdot] &= \tilde{a}_g \cdot A_g \cdot e^{j\boldsymbol{\theta}_g} \\ &= (a_{gI} + ja_{gQ}) \cdot A_g \cdot (cos\boldsymbol{\theta}_g + jsin\boldsymbol{\theta}_g) \\ &= a_{gI}A_g cos\boldsymbol{\theta}_g - a_{gQ}A_g sin\boldsymbol{\theta}_g \\ &\quad + j(a_{gI}A_g sin\boldsymbol{\theta}_g + a_{gQ}A_g cos\boldsymbol{\theta}_g) \end{aligned} \tag{2}$$

where $F_g[\cdot] = F_{gI}[\cdot] + jF_{gQ}[\cdot]$, $F_{gI}[\cdot]$ and $F_{gQ}[\cdot]$ represent the real and imaginary parts of $F_g[\cdot]$:

$$F_{gI}[\cdot] = a_{gI}A_g \cos\boldsymbol{\theta}_g - a_{gQ}A_g \sin\boldsymbol{\theta}_g \tag{3}$$

$$F_{gQ}[\cdot] = a_{gQ}A_g \cos\boldsymbol{\theta}_g + a_{gI}A_g \sin\boldsymbol{\theta}_g \tag{4}$$

In order to realize the mathematical relationship in (2) with a neural network architecture, a partial network for the gth basis function is designed as Fig. 1.

In the first stage, the magnitudes of the input signal, $|\tilde{x}(n)|$, pass through the gth hidden neuron with activation function

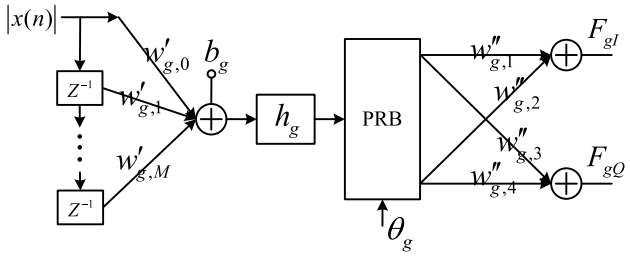


FIGURE 1. Proposed vector decomposed sub-network.

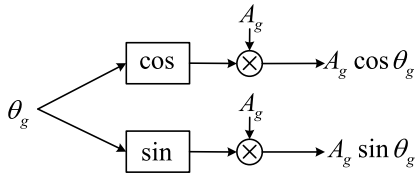


FIGURE 2. Phase recovery block architecture.

h_g and the neuron output A_g is obtained as follows:

$$A_g = h_g\left[\sum_{m=0}^{M} w'_{g,m}\,|\tilde{x}(n-m)| + b_g\right] \tag{5}$$

where $(w'_{g,0}, w'_{g,1}, \cdots, w'_{g,M})$ are the weighting coefficients from the input layer to the gth hidden neuron and b_g is the bias before the hidden neuron.

Secondly, the phase information is recovered in the phase recovery block (PRB), where A_g is multiplied with phase information $cos\boldsymbol{\theta}_g$ and $sin\boldsymbol{\theta}_g$, respectively. $\boldsymbol{\theta}_g$ is the input phase information to the PRB. PRB is a dual-input and dual-output block as shown in Fig. 2, where $\boldsymbol{\theta}_g$ and A_g are the inputs to the block. Send $\boldsymbol{\theta}_g$ into the block to calculate its sine and cosine values, then $cos\boldsymbol{\theta}_g$ and $sin\boldsymbol{\theta}_g$ are multiplied with the output of the hidden neuron A_g, where $A_g\cos\boldsymbol{\theta}_g$ and $A_g\sin\boldsymbol{\theta}_g$ are obtained at the upper output and lower output of PRB.

Finally, take $(A_g sin\boldsymbol{\theta}_g, A_g cos\boldsymbol{\theta}_g)$ and (Y_{gI}, Y_{gQ}) as the input and output of the last fully-connected layer to finish the sub-network construction. $(w''_{g,1}, w''_{g,2}, w''_{g,3}, w''_{g,4})$ are the weighting coefficients in the last fully-connected layer. According to Fig. 1, the output of the sub-network are as follows:

$$Y_{gI} = w''_{g,1}A_g\cos\boldsymbol{\theta}_g + w''_{g,2}A_g\sin\boldsymbol{\theta}_g \tag{6}$$

$$Y_{gQ} = w''_{g,3}A_g\cos\boldsymbol{\theta}_g + w''_{g,4}A_g\sin\boldsymbol{\theta}_g \tag{7}$$

And it is not difficult to find that the ideal corresponding relationship among (3), (4), (6) and (7) is: $w''_{g,1} = w''_{g,4} = a_{gI}$, $w''_{g,2} = -a_{gQ}$ and $w''_{g,3} = a_{gQ}$. It is worth mentioning that there is no bias in the output neuron layer, otherwise, it will introduce nonphysical contribution [23], [24].

B. OBTAIN LARGE NETWORK FROM THE EXTRACTED SUB-NETWORK

When we finished the above sub-network construction for the gth basis function, the sub-network can be expanded to a large network with G hidden neurons in the hidden layer, as is depicted in Fig. 3, which corresponds to G basis functions in (1). $I_{out}(n)$ and $Q_{out}(n)$ are the real and imaginary parts of $\tilde{y}(n)$. For ease of illustration, the G hidden neurons are renumbered into different groups, where there are $(M+1)$ hidden neurons in each group: $(h_{i,0}, h_{i,1}, \cdots, h_{i,M})$ represent the ith group with $(M+1)$ hidden neurons.

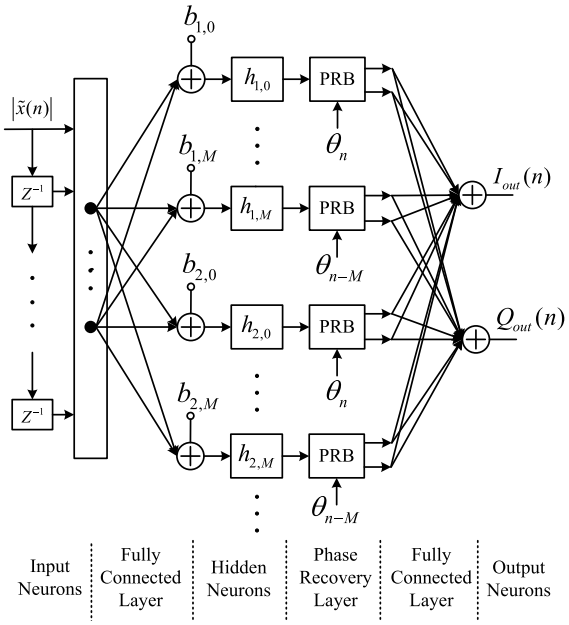


FIGURE 3. The expanded large network with G hidden neurons.

In order to satisfy the "first zone constraint", in this paper, the phase of each hidden neuron's output is finally restored to that of the original input signal in a cyclic principle: the phases of each hidden neuron group outputs $(A_{i,0}, A_{i,1}, \cdots, A_{i,M})$ are restored with original phase information with $(e^{j\boldsymbol{\theta}_n}, e^{j\boldsymbol{\theta}_{n-1}}, \cdots, e^{j\boldsymbol{\theta}_{n-M}})$, respectively. In this manner, the phase information for other $(G-M-1)$ neurons can be also recovered with the same manner until to the last hidden neuron $h_{u,v}$ in the hidden layer, where $(M+1)\times u + v = G$.

As is shown in Fig. 3: First, the magnitudes of the input signals pass the first fully-connected layer and the hidden neurons to conduct the nonlinear operation for the envelop signals; Second, the outputs of the hidden neurons are multiplied with the phase information in the above mentioned cyclic principle to realize the phase recovery. Finally, take the output of the phase recovery layer and $(I_{out}(n), Q_{out}(n))$ as the input and the output of the last fully-connected layer to complete the large network construction. For ease of drawing picture, the weighting coefficients in the two fully-connected layer are omitted in Fig. 3, so are the model architecture pictures in the remainder of this article.

C. SIMPLIFY THE PROPOSED LARGE NETWORK

After the corresponding neural network architecture of (1) was obtained, it was found that the proposed neural network model can be simplified further as follows:

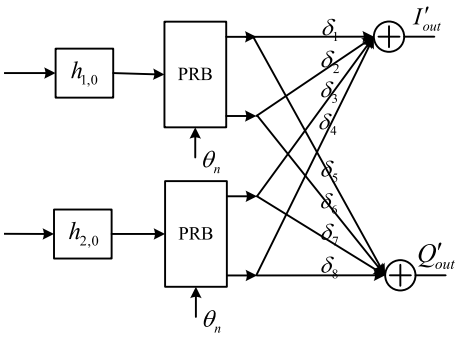


FIGURE 4. A sub-network with the same recovery phase information.

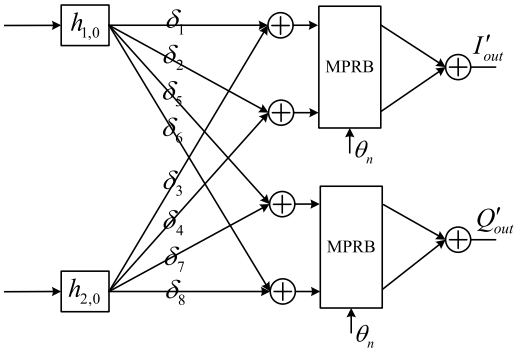


FIGURE 5. The equivalent neural network architecture of (10) and (11).

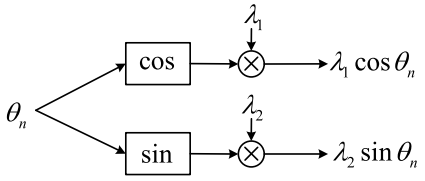


FIGURE 6. Modified phase recovery block architecture.

In order to illustrate the simplification procedure clearly, take a sub-network of Fig. 3, where only two hidden neurons $h_{1,0}$ and $h_{2,0}$ with the same recovered phase information are considered, as an example to explain this simplification. As is shown in Fig. 4 (the input layer and the first fully-connected layer are omitted here), $(\delta_1, \delta_2, \cdots, \delta_8)$ are the weighting coefficients in their corresponding paths. I'_{out} and Q'_{out} are the in-phase and quadrature parts of the sub-network output. If the output of $h_{1,0}$ and $h_{2,0}$ are defined as $A_{1,0}$ and $A_{2,0}$, the corresponding mathematical expressions of Fig. 4 are as follows:

$$\begin{aligned} I'_{out} &= \boldsymbol{\delta}_1 A_{1,0}\cos\boldsymbol{\theta}_n + \boldsymbol{\delta}_2 A_{1,0}\sin\boldsymbol{\theta}_n \\ &\quad + \boldsymbol{\delta}_3 A_{2,0}\cos\boldsymbol{\theta}_n + \boldsymbol{\delta}_4 A_{2,0}\sin\boldsymbol{\theta}_n \qquad (8) \\ Q'_{out} &= \boldsymbol{\delta}_5 A_{1,0}\cos\boldsymbol{\theta}_n + \boldsymbol{\delta}_6 A_{1,0}\sin\boldsymbol{\theta}_n \\ &\quad + \boldsymbol{\delta}_7 A_{2,0}\cos\boldsymbol{\theta}_n + \boldsymbol{\delta}_8 A_{2,0}\sin\boldsymbol{\theta}_n \qquad (9) \end{aligned}$$

After combining the similar terms with the same $\cos\boldsymbol{\theta}_n$ and $\sin\boldsymbol{\theta}_n$ factor, (8) and (9) can be rewritten as (10) and (11), the corresponding network architecture is depicted in Fig. 5.

$$\begin{aligned} I'_{out} &= (\boldsymbol{\delta}_1 A_{1,0} + \boldsymbol{\delta}_3 A_{2,0})\cos\boldsymbol{\theta}_n \\ &\quad + (\boldsymbol{\delta}_2 A_{1,0} + \boldsymbol{\delta}_4 A_{2,0})\sin\boldsymbol{\theta}_n \qquad (10) \\ Q'_{out} &= (\boldsymbol{\delta}_5 A_{1,0} + \boldsymbol{\delta}_7 A_{2,0})\cos\boldsymbol{\theta}_n \\ &\quad + (\boldsymbol{\delta}_6 A_{1,0} + \boldsymbol{\delta}_8 A_{2,0})\sin\boldsymbol{\theta}_n \qquad (11) \end{aligned}$$

Define $\lambda_1 = \boldsymbol{\delta}_1 A_{1,0} + \boldsymbol{\delta}_3 A_{2,0}$ and $\lambda_2 = \boldsymbol{\delta}_2 A_{1,0} + \boldsymbol{\delta}_4 A_{2,0}$, we take the first modified phase recovery block (MPRB) in Fig. 5 as an example to illustrate MPRB architecture: MPRB is a three-input double-output block as shown in Fig. 6, send $\boldsymbol{\theta}_g$ into the block to calculate its sine and cosine values, then $cos\boldsymbol{\theta}_g$ and $sin\boldsymbol{\theta}_g$ are multiplied with λ_1 and λ_2, respectively. Thus, $A_g\cos\boldsymbol{\theta}_g$ and $A_g\sin\boldsymbol{\theta}_g$ are obtained at the upper output and lower output of MPRB. As is shown in Fig. 5, there are two MPRBs in the sub-network for the phase information recovery: one for the in-phase output and the other for the quadrature output.

As the corresponding complex mathematical expression of Fig.3 is (1), similar to the above operation process, split the output $\tilde{y}(n)$ in (1) into in-phase and quadrature parts and combine those terms including the same $\cos\boldsymbol{\theta}_{n-m}$ and $\sin\boldsymbol{\theta}_{n-m}$ factors, it yields:

$$\begin{aligned} I_{out}(n) &= \sum_{m=1}^{M} \boldsymbol{\alpha}_{Im}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)\cos\boldsymbol{\theta}_{n-m} \\ &\quad + \sum_{m=1}^{M} \boldsymbol{\beta}_{Im}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)\sin\boldsymbol{\theta}_{n-m} \qquad (12) \\ Q_{out}(n) &= \sum_{m=1}^{M} \boldsymbol{\alpha}_{Qm}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)\cos\boldsymbol{\theta}_{n-m} \\ &\quad + \sum_{m=1}^{M} \boldsymbol{\beta}_{Qm}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)\sin\boldsymbol{\theta}_{n-m} \qquad (13) \end{aligned}$$

where $\boldsymbol{\alpha}_{Im}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)$ and $\boldsymbol{\beta}_{Im}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)$ are the linear summation of the hidden neurons output $(A_{1,m}, A_{2,m}, A_{3,m}\cdots)$ after the same terms $\cos\boldsymbol{\theta}_{n-m}$ and $\sin\boldsymbol{\theta}_{n-m}$ combination and extraction for the in-phase output; $\boldsymbol{\alpha}_{Qm}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)$ and $\boldsymbol{\beta}_{Qm}(A_{1,m}, A_{2,m}, A_{3,m}, \cdots)$ are the linear summation of hidden neurons output $(A_1, A_2, \cdots A_G)$ after the same terms $\cos\boldsymbol{\theta}_{n-m}$ and $\sin\boldsymbol{\theta}_{n-m}$ combination and extraction for the quadrature output.

Thus, after those terms combination with the same phase factor $\cos\boldsymbol{\theta}_{n-m}$ and $\sin\boldsymbol{\theta}_{n-m}$, the hidden neurons in the hidden layer of Fig. 3 can be rearranged into $(M+1)$ group and each group has their corresponding phase factor. The lth $(l = 1, \cdots, M+1)$ hidden neuron group $(h_{1,l-1}, h_{2,l-1}, h_{3,l-1}, \cdots)$ are recovered with the same phase information $\boldsymbol{\theta}_{n-l+1}$. In this way, the neural network architecture in Fig.3 can be simplified as that in Fig. 7. By rearranging the hidden neurons into $(M+1)$ groups to recover the phases, the number of multiplication for phase information recovery

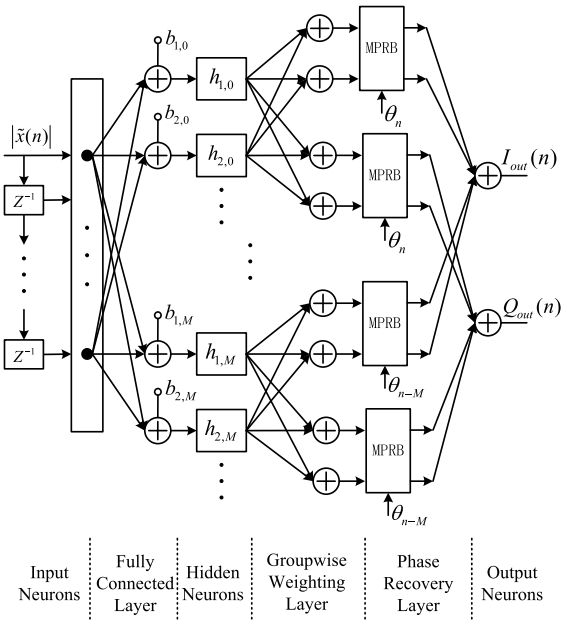


FIGURE 7. The simplified neural network architecture of Fig. 3.

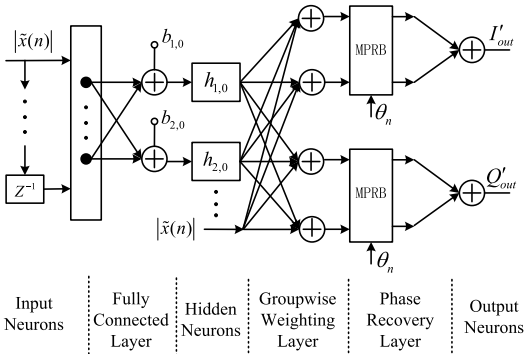


FIGURE 8. The first hidden neuron group with linear term neuron $|\tilde{x}(n)|$.

can be reduced from $2G$ to $4(M+1)$ (Generally, the number of hidden neurons G would be much larger than that of memory depth M). With the increase of the hidden neurons, the multiplication number used for phase recovery remains unchanged.

D. ADD LINEAR TERM NEURONS

Generally, in the conventional RF PA models [3], both linear and nonlinear terms should be included. Although neural network is a general nonlinear fitting method, in RF PA modeling, it was found that adding linear term neurons could further improve the model performance. In this part, the magnitudes of the input signals with different memory depths are augmented into their corresponding phase memory depth group as the input of the groupwise weighting layer to realize linear terms implementation. For example, as is shown in Fig. 8, without nonlinear operations $|\tilde{x}(n)|$ are augmented as the input of the groupwise weighting layer of the first hidden neuron group. In exactly the same way, other M linear term neurons can be implemented in the proposed MLP network.

TABLE 1. NMSE performance comparison of adding linear term neurons.

Activation Function	Proposed Method without Linear Term Neurons (dB)	Proposed Method with Linear Term Neurons (dB)
ABS	-38.07	-38.70
ReLU	-37.20	-37.64
Sigmoid	-36.66	-37.59
Tanh	-36.86	-37.47

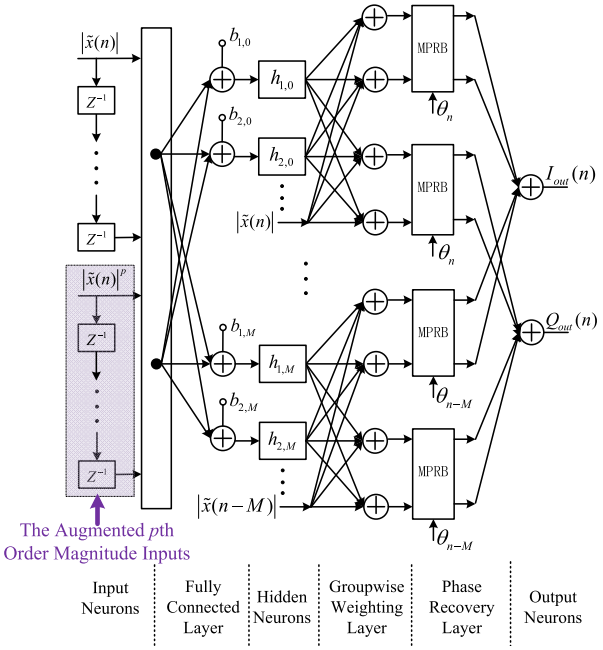


FIGURE 9. Proposed VDTDNN model and AVDTDNN model architectures.

Finally, it yields the proposed vector decomposed time-delay neural network (VDTDNN) model, as is depicted in Fig. 9. TABLE 1 shows that the modeling capacity of the proposed VDTDNN model with different activation functions can be improved by adding linear term neurons at the cost of very little computational complexity, where the number of hidden neurons for TABLE 1 is 30 and the other settings are the same as those in PART A of Section V.

Since the three constraints mentioned at the beginning of this section can be satisfied simultaneously, the proposed DPD model, theoretically, can achieve better performance compared with the existing IQ-mapping-IQ NN DPD models. In addition, since the complex baseband input signals are decomposed into magnitude and phase parts, where only the magnitudes are conducted nonlinear operations and phase information is recovered in the phase recovery layer with linear weightings, that is why the proposed model is named as vector decomposed time-delay neural network (VDTDNN) behavioral model.

III. AUGMENTED VECTOR DECOMPOSED TIME-DELAY NEURAL NETWORK

Recently, an augmented real-valued time-delay neural network (ARVTDNN) [26] has been proposed to improve

the performance of real-valued time-delay neural network (RVTDNN), which shows that adding several envelop-dependent terms into the input vector could improve the modeling capability of MLP neural networks, and the experimental results shows that the modeling capability of ARVTDNN are superior to RVTDNN and memory polynomial models. In this paper, the high order envelop-dependent terms are also added into the input vector, and it yields the augmented vector decomposed real-valued time-delay neural network (AVDTDNN) model as shown in Fig. 9, where the shaded part represents the augmented pth order magnitude inputs. The experimental results prove that the proposed AVDTDNN model can achieve better performance and with lower computational complexity compared with the existing ARVTDNN model. The detailed computational complexity analysis and tested performance comparison for AVDTDNN and ARVTDNN models can be seen in PART B of Section IV and PART B of Section V, respectively.

IV. COMPUTATIONAL COMPLEXITY COMPARISON ANALYSIS

In this part, the complexity comparison analysis reveals that whether envelop-dependent terms are augmented into the input vector or not, the computational complexity of the proposed VDTDNN and AVDTDNN are superior to the existing IQ-mapping-IQ based RVTDNN and ARVTDNN, respectively. It is analyzed in details as follows:

A. THE COMPLEXITY COMPARISON OF RVTDNN AND VDTDNN

When there are no high order envelop-dependent terms adding into the input vector, the complexity comparison of RVTDNN and VDTDNN are as follows:

Given that the number of nonlinear operation hidden layer and its hidden neurons are 1 and G. Memory depth is M. As the weighting factors in the phase recovery layer are known, so only the coefficients in the first fully connected layer and groupwise weighting layer should be calculated with back propagation. The total coefficients to be updated in the RVTDNN is:

$$N_0 = (2 \times (M + 1) + 1) \times G + (G + 1) \times 2 \tag{14}$$

While the number of coefficients to be updated in the VDTDNN is:

$$N_1 = (M + 1 + 1) \times G + 4 \times (G + M + 1) \tag{15}$$

And it yields:

$$N_0 - N_1 = M \times G - G - 2 \times M - 2 \tag{16}$$

The detailed values of $N_0 - N_1$ and its corresponding memory depth M and the number of hidden neurons are shown in Fig. 10, once $M \geq 2$, the number of $N_0 - N_1$ would be larger than 0. And generally, in order to offer a relatively good modeling performance, we have $M \geq 2$ and $G \geq 10$ (especially for wideband signals), it means that the computational complexity of the proposed VDTDNN model is lower

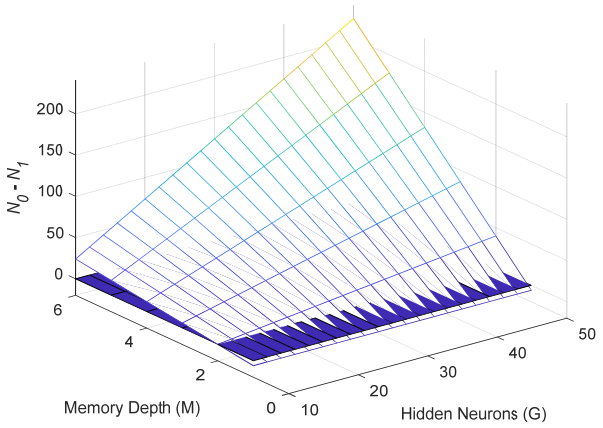


FIGURE 10. The relationship of (16).

TABLE 2. Performance comparison of VDTDNN and RVTDNN models.

Model Type	Total Coefficients	NMSE (dB)	ACPR (dB)
Without DPD	N/A	-24.41	-29.8/-29.16
RVTDNN	392	-34.65	-41.05/-40.03
VDTDNN	320	-38.60	-46.27/-46.28

than that of the existing RVTDNN model. Take $M = 4$ and $G = 30$ as an example, the total coefficients to be updated in RVTDNN and VDTDNN models are 392 and 320, respectively, as was listed in TABLE 2.

B. THE COMPLEXITY COMPARISON OF ARVTDNN AND AVDTDNN

The experimental results in [26] show that adding $(|\tilde{x}(n-m)|, |\tilde{x}(n-m)|^2, |\tilde{x}(n-m)|^3)$ or $(|\tilde{x}(n-m)|, |\tilde{x}(n-m)|^3, |\tilde{x}(n-m)|^5)$ terms into the input vector could offer the least computational complexity with relatively good performance. In this paper, take it as an example by adding $(|\tilde{x}(n-m)|, |\tilde{x}(n-m)|^2, |\tilde{x}(n-m)|^3)$ into the input to compare the complexity between ARVTDNN and AVDTDNN. As the input to the proposed VDTVNN are linear magnitudes instead of in-phase and quadrature parts of complex baseband signals, thus, only $(|\tilde{x}(n-m)|^2, |\tilde{x}(n-m)|^3)$ terms are needed to add into the input vector to construct AVDTDNN model, it yields:

$$N_2 = (2 \times (M+1) + 3 \times (M+1) + 1) \times G + (G+1) \times 2 \tag{17}$$

$$N_3 = (M + 1 + 2 \times (M+1) + 1) \times G + 4 \times (G + M + 1) \tag{18}$$

And it yields:

$$N_2 - N_3 = 2 \times M \times G - 4 \times M - 2 \tag{19}$$

where N_2 and N_3 are the total coefficients in the ARVTDNN and AVDTDNN to be extracted, respectively. The detailed relationship of (19) is depicted in Fig. 11, which indicates that the computational complexity of AVDTDNN is lower

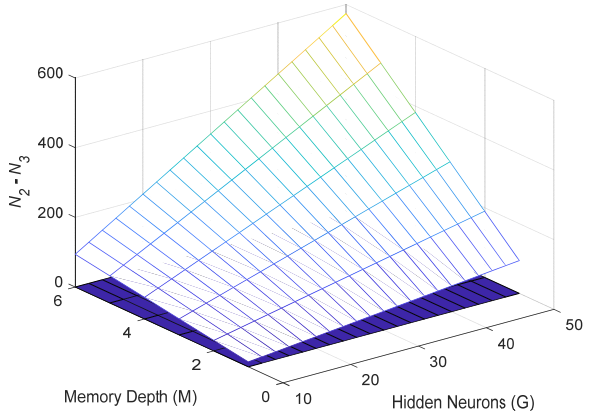


FIGURE 11. The relationship of (19).

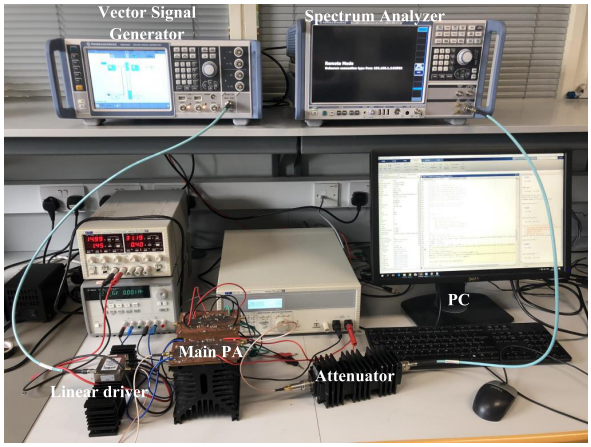


FIGURE 12. The photograph of the test bench.

than that of ARVTDNN. Take $M = 4$ and $G = 30$ as an example, the total coefficients to be updated in ARVTDNN and AVDTDNN models are 842 and 620, respectively.

V. EXPERIMENTAL VALIDATION

In order to compare the performance of the proposed models and the existing IQ-mapping-IQ based models, various experimental tests were conducted. The photograph of the test bench is shown in Fig. 12, which consists of a personal computer (PC) with MATLAB and pytorch softwares, a Vector Signal Generator (SMW2000A) from Rohde and Schwarz, a linear driver amplifier PA, a house-designed wideband GaN PA, a -30 dB RF attenuator and a Spectrum Analyzer (FSW50) from Rohde and Schwarz. The flow diagram of the modeling procedure with indirect learning architecture (ILA) is depicted in Fig. 13. First the upconverted RF signals were generated by the Vector Signal Generator under the control of PC. Second, after a linear driver PA, the RF signals were amplified with the GaN wideband PA operated at 3.75GHz, which is possible for 5G application. Third, the attenuated PA outputs were captured by the Spectrum Analyzer and sent into the PC. Finally, after time alignment with MATLAB, according to classical indirect learning architecture, the baseband input and output of the RF PA were sent into

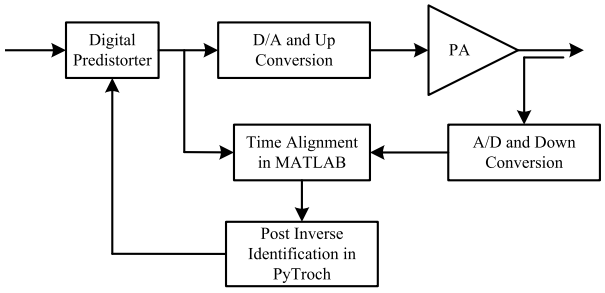


FIGURE 13. The flow diagram of the modeling procedure with ILA.

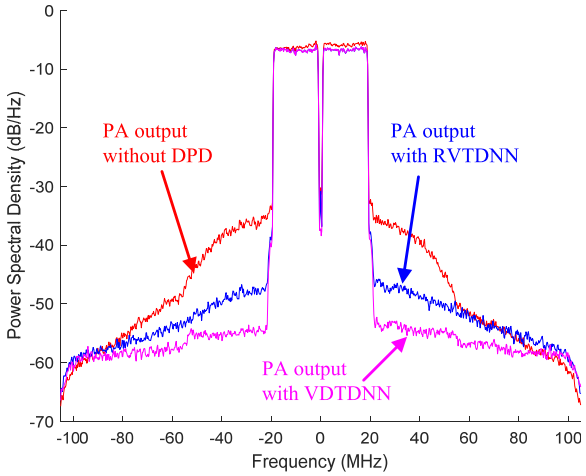


FIGURE 14. The power spectral density comparison.

PyTorch and used as output and input of the DPD model to extract model coefficients (Note: it is more convenient and flexible to construct neural network models with pytorch compared with MATLAB, thus, the neural network models were implemented in PyTroch with python language in this paper).

A. THE PERFORMANCE COMPARISON OF RVTDNN AND VDTDNN

Take one hidden layer as an example to compare the performance of the proposed VDTDNN model and RVTDNN model, where memory depth is 4, the number of hidden neuron is 30, and number of iteration is 150. The bandwidth of the tested signal is 40 MHz OFDM signal and its PAPR is 6.9 dB. Adaptive Moment (Adam) optimizer was used to update the coefficients [27]. The step sizes for the first 130 iterations and the last 20 iterations are 0.01 and 0.001, respectively. Around 70,000 samples were recorded with the sampling rate at 368.64 MSPS. The first 40,000 samples were used to extract model coefficients and the remaining samples were used for performance evaluation. The RF PA output power was 30.3dBm. Both VDTDNN and RVTDNN models keep the same settings as mentioned above.

When the activation function in the first hidden layer is selected as absolute value function (ABS), the detailed normalized mean square error (NMSE) and adjacent channel

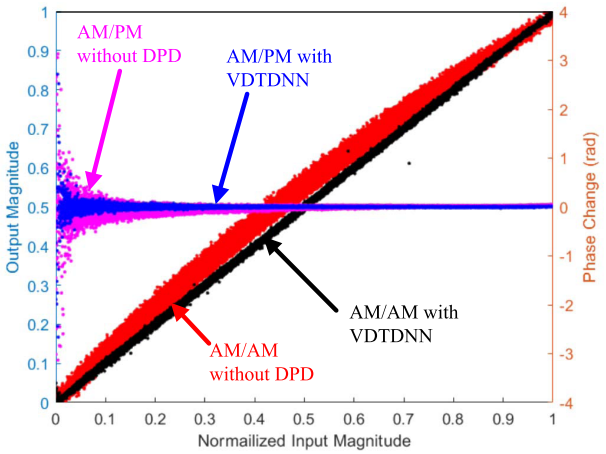


FIGURE 15. AM-AM and AM-PM plots with and without DPD for a 40-MHz OFDM signal.

TABLE 3. Performance comparison with different activation functions.

Type of Activation Functions	RVTDNN		VDTDNN	
	NMSE (dB)	ACPR (dB) ($\pm 40MHz$)	NMSE (dB)	ACPR (dB) ($\pm 40MHz$)
Without DPD	-24.41(-29.8/-29.16)			
ABS	-35.53	-41.97/-40.68	-38.41	-46.88/-46.37
ReLU	-35.78	-43.72/-42.11	-37.58	-45.50/-45.63
Sigmoid	-35.80	-41.57/-43.07	-37.66	-45.62/-44.76
Tanh	-35.73	-42.31/-42.58	-36.78	-43.73/-43.59

power ratio (ACPR) are listed in TABLE 2. The corresponding Power Spectral Density (PSD) comparison is illustrated in Fig. 14. According to TABLE 2 and Fig. 14, compared with RVTDNN model, NMSE and ACPR with VDTDNN model can be improved by 3.5dB and 5dB respectively and with less coefficients. And the AM-AM and AM-PM characteristics without DPD and with VDTDNN model DPD are shown in Fig. 15.

For further comparison, four commonly used activation functions, ABS, rectified linear unit (ReLU), Sigmoid and hyperbolic tangent (Tanh), were tested to compare the performance of the existing RVTDNN model and the proposed VDTDNN model as shown in Table 3. It is clear that with any kinds of activation function, the performance of the proposed model are superior to the existing state-of-the-art RVTDNN model.

B. THE PERFORMANCE COMPARISON OF ARVTDNN AND AVDTDNN

To further compare the performance of the proposed vector decomposed based models and the state-of-the-art MLP based models, a 60MHz OFDM signal was employed to compare performance of RVTDNN, ARVTDNN, VDTDNN and AVDTDNN models, where the PA output power is 32.6dBm and the other settings are the same as those in part A of this section. TABLE 4 gives the NMSE, ACPR and total coefficients results. Fig. 16 shows PA output power spectrum density curves without DPD and with various DPD models. When there are no other envelop-based terms added, the novel VDTDNN can approximately further improve NMSE 3.4dB and ACPR 4.5 dB compared with RVTDNN model, respectively; when the envelop-based terms ($|\tilde{x}(n-m)|$, $|\tilde{x}(n-m)|^2$, $|\tilde{x}(n-m)|^3$) were added into the input, the proposed AVDTDNN model achieved the best performance but with significantly reduced coefficients, compared with the state-of-the-art ARVTDNN model.

TABLE 4. Performance comparison of the mentioned models.

Type of Model	NMSE (dB)	ACPR(dB) ($\pm 60MHz$)	Number of Coefficients
Without DPD	-19.75	-25.04/-24.21	N/A
VDTDNN	-36.12	-44.59/-43.06	320
AVDTDNN	-38.50	-48.95/-49.09	620
RVTDNN	-32.75	-39.91/-39.47	392
ARVTDNN	-36.79	-46.11/-45.92	842

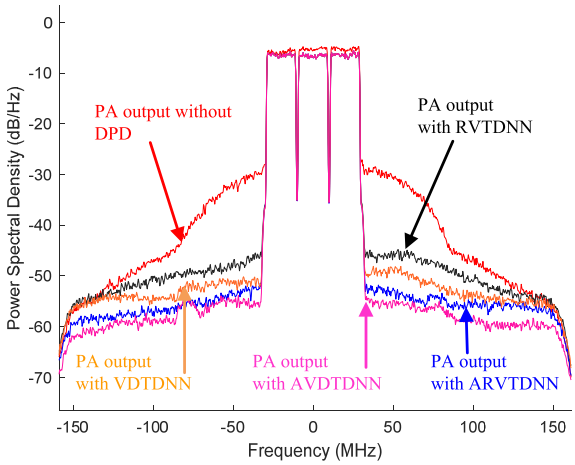


FIGURE 16. The power spectral density comparison.

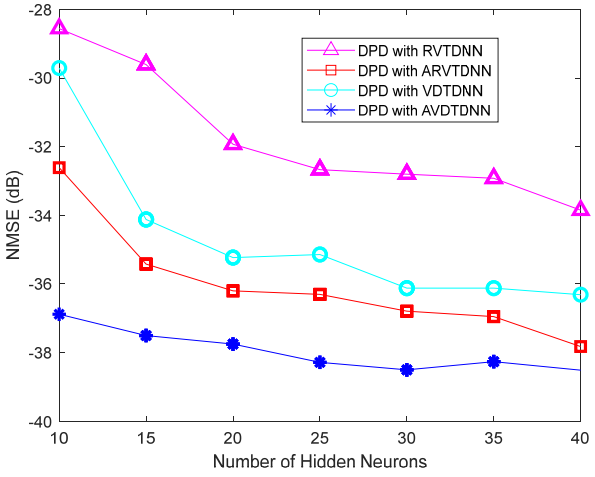


FIGURE 17. NMSE comparison with varied numbers of hidden neuron.

In order to consummate the experimental comparison, the number of hidden neurons are then changed to evaluate the performance. The number of hidden neurons varied from 10 to 40 at the interval of 5. The activation function is selected as ABS. As can be seen from Fig. 17, without envelop-based terms augmented, compared with RVTDNN, the NMSE can be improved about 4dB with VDTDNN model; with envelop-based terms in the input, the NMSE can be improved 2dB with AVDTDNN model compared with the existing ARVTDNN model. Consequently, whether other envelop-based terms are augmented into the input vector or not, the proposed VDTDNN and AVDTDNN models are superior to the state-of-the-art RVTDNN and ARVTDNN models.

VI. CONCLUSION

In this study, two novel vector-decomposition based MLP neural network DPD models are proposed. Different from the state-of-the-art neural network models splitting both input and output into IQ parts to construct models, the proposed models are in more accordance with the physical mechanisms of RF PA, where only the magnitudes of input signals are conducted nonlinear operations, and the phase information is then recovered with linear weighting operations. The theoretical analysis and various experimental results show that the proposed VDTDNN and AVDTDNN can achieve better performance and with significantly lower computational complexity, compared with the existing neural network DPD models.

REFERENCES

[1] A. Gupta and E. R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

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

[4] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

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

[6] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, "Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018.

[7] O. Hammi and F. M. Ghannouchi, "Twin nonlinear two-box models for power amplifiers and transmitters exhibiting memory effects with application to digital predistortion," *IEEE Microw. Compon. Lett.*, vol. 19, no. 8, pp. 530–532, Aug. 2009.

[8] S. Boumaiza, J. Li, M. Jaidane-Saidane, and F. M. Ghannouchi, "Adaptive digital/RF predistortion using a nonuniform LUT indexing function with built-in dependence on the amplifier nonlinearity," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 12, pp. 2670–2677, Dec. 2004.

[9] W. Cao and A. Zhu, "A modified decomposed vector rotation-based behavioral model with efficient hardware implementation for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2443–2452, Jul. 2017.

[10] M. Younes, O. Hammi, A. Kwan, and F. M. Ghannouchi, "An accurate complexity-reduced 'PLUME' model for behavioral modeling and digital predistortion of RF power amplifiers," *IEEE Trans. Ind. Electron.*, vol. 58, no. 4, pp. 1397–1405, Apr. 2011.

[11] R. Hongyo, Y. Egashira, T. M. Hone, and K. Yamaguchi, "Deep neural network-based digital predistorter for Doherty power amplifiers," *IEEE Microw. Wireless Compon. Lett.*, vol. 29, no. 2, pp. 146–148, Feb. 2019.

[12] A. Zhu, "Behavioral modeling for digital predistortion of RF power amplifiers: From Volterra series to CPWL functions," in *Proc. IEEE Top. Conf. Power Amplif. Wireless Radio Appl. (PAWR)*, Jan. 2016, pp. 1–4.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[14] A. Ahmed, E. R. Srinidhi, and G. Kompa, "Efficient PA modeling using neural network and measurement setup for memory effect characterization in the power device," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2005, pp. 473–476.

[15] H. Leung and S. Haykin, "The complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 2101–2104, Sep. 1991.

[16] 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.

[17] 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.

[18] 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.

[19] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.

[20] 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.

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

[22] 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.

[23] E. G. Lima, T. R. Cunha, H. M. Teixeira, M. Pirola, and J. C. Pedro, "Baseband derived Volterra series for power amplifier modeling," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2009, pp. 1361–1364.

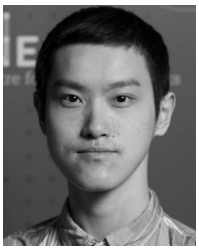
[24] E. G. Lima, T. R. Cunha, and J. C. Pedro, "A physically meaningful neural network behavioral model for wireless transmitters exhibiting PM–AM/PM–PM distortions," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3512–3521, Dec. 2011.

[25] F. M. Ghannouchi, O. Hammi, and M. Helaoui, *Behavioral Modeling and Predistortion of Wideband Wireless Transmitters*. New York, NY, USA: Wiley, 2015.

[26] 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, Jun. 2019.

[27] J. Ba and D. P. Kingma, "Adam: A method for stochastic optimization," in *Proc. 32nd Int. Conf. Learn. Represent.*, 2015, pp. 127–142.

YIKANG ZHANG received the B.E. degree in electronic science and technology from the China University of Mining and Technology, Xuzhou, China, in 2015. He is currently pursuing the Ph.D. degree in electronic engineering with the University of Science and Technology of China, Hefei, China. From 2018 to 2019, he joined the IoE2 Laboratory, School of Electrical and Electronic and Engineering, University College Dublin, Dublin, Ireland, as a Visiting Ph.D. Student. His research interests include digital predistortion linearization, nonlinear system identification algorithms, and machine learning.



YUE LI (S'17) received the B.E. degree in information engineering from Southeast University, Nanjing, China, in 2016. He is currently pursuing the Ph.D. degree with University College Dublin (UCD), Dublin, Ireland, where he is also with the RF and Microwave Research Group. His current research interests include behavioral modeling and digital predistortion for radio-frequency (RF) power amplifiers.

FALIN LIU was born in Xingtai, Hebei, China, in 1963. He received the B.E. degree from Tsinghua University (THU), Beijing, China, in 1985, and the M.E. and Ph.D. degrees from the University of Science and Technology of China (USTC), Hefei, China, in 1988 and 2004, respectively, all in electronic engineering. Since 1988, he has been with the Department of Electronic Engineering and Information Science, USTC, where he is currently a Full Professor. From 1997 to 1998, he was a Visiting Scholar with Tohoku University, Japan. He has published over 80 technical papers in refereed journals and international conferences. His current research interests include computational electromagnetics, microwave circuits and systems, microwave imaging, and satellite communications. He is also a Senior Member of the Chinese Institute of Electronics. He serves on the Editorial Board of the *Journal of Microwaves* (in Chinese) and the *Journal of Radars* (in Chinese).

ANDING ZHU (S'00–M'04–SM'12) received the B.E. degree in telecommunication engineering from North China Electric Power University, Baoding, China, in 1997, the M.E. degree in computer applications from the Beijing University of Posts and Telecommunications, Beijing, China, in 2000, and the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2004.

He is currently a Professor with the School of Electrical and Electronic Engineering, UCD. He has published over 100 peer-reviewed journals and conference papers. His research interests include high-frequency nonlinear system modeling and device characterization techniques with a particular emphasis on behavioral modeling and linearization of radio-frequency (RF) power amplifiers for wireless communications, and high-efficiency power amplifier design, wireless transmitter architectures, digital signal processing, and nonlinear system identification algorithms.

Dr. Zhu served as the Secretary of the IEEE MTT-S ADCOM, in 2018. He is also an elected member of MTT-S ADCOM, an Editorial Board Member of the IEEE Future Networks Tech Focus, and the Vice Chair of the Publications and Electronic Information Committees. He was the General Chair of the 2018 IEEE MTT-S International Microwave Workshop Series on 5G Hardware and System Technologies (IMWS-5G). He is also an Associate Editor of *IEEE Microwave Magazine*.