Vector Decomposed Long Short-Term Memory Model for Behavioral Modeling and Digital Predistortion for Wideband RF Power Amplifiers

HONGMIN LI [ID], YIKANG ZHANG [ID], GANG LI [ID], AND FALIN LIU [ID]

[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

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

ABSTRACT This paper proposes two novel vector decomposed neural network models for behavioral modeling and digital predistortion (DPD) of radio-frequency (RF) power amplifiers (PAs): vector decomposed long short-term memory (VDLSTM) model and simplified vector decomposed long short-term memory (SVDLSTM) model. The proposed VDLSTM model is a variant of the classic long short-term memory (LSTM) model that can model long-term memory effects. To comply with the physical mechanism of RF PAs, VDLSTM model only conducts nonlinear operations on the magnitudes of the input signals, while the phase information is recovered by linear weighting operations on the output of the LSTM cell. Furthermore, this study modifies the LSTM cell by adding phase recovery operations inside the cell and replacing the original hidden state with the output magnitudes that are recovered with phase information. With the modified LSTM cell, a low-complexity SVDLSTM model is proposed. The experiment results show that the proposed VDLSTM model can achieve better linearization performance compared with the state-of-the-art models when linearizing PAs with wideband inputs. Besides, in wideband scenarios, SVDLSTM model with much fewer parameters can present comparable linearzation performance compared to VDLSTM model.

INDEX TERMS Nonlinear power amplifier, behavioral modeling, digital predistortion, neural network, long short-term memory, vector decomposed.

I. INTRODUCTION

In the advanced fifth-generation (5G) wireless communication systems, carrier frequencies and signal bandwidths increase significantly driven by the high-capacity demand [1], [2]. Wideband signals modulated at high carrier frequencies have high peak-to-average ratio (PAPR) and encounter a conflict between linearity and efficiency of radio frequency (RF) power amplifiers (PAs). Due to the inherent nonlinear characteristics, RF PAs working in a high-efficiency state introduce severe nonlinearities, which results in bit error rate (BER) deterioration and adjacent channel interference [3]. Lots of linearization methods have been developed to balance the trade-off between linearity and efficiency, such as

feed-forward technology, analog predistortion and digital predistortion (DPD) [4]. Among them, DPD is generally believed to be the most promising linearization technology for its flexibility and high performance. The key of DPD is to extract an inverse behavioral model called digital predistorter and place it before the PA that needs to be linearized. The digital predistorter is implemented in digital domain. Then, the cascade of the digital predistorter and the nonlinear RF PA will be a linear system [5]. By selecting the proper behavioral models for the digital predistorter, DPD can effectively compensate the nonlinearities of the RF PAs.

Various DPD models have been proposed to compensate the nonlinearities of RF PAs [5]–[10]. Volterra-based models such as memory polynomial (MP) model [5], generalized memory polynomial (GMP) model [6] are the most widely used ones. However, the basis functions of the Volterra-based

---

The associate editor coordinating the review of this manuscript and approving it for publication was Juntao Fei [ID].

models are polynomials, which leads to high correlation among different basis functions [11]. This characteristic of Volterra-based models would limit the performance even with a large amount of basis functions, and may deteriorate numerical instability [12]. Based on canonical piecewise-linear (CPWL) functions, another state-of-the-art model, decomposed vector rotation (DVR) model was proposed to model the PAs with strong nonlinearities such as envelope tracking (ET) PAs [9]. DVR model is an effective supplement to the Volterra-based models. However, since signals employed in 5G communication systems have wider bandwidth and higher PAPR, both the static nonlinearity and memory effects of RF PAs become much more severe and complex [13], which results in the limited linearization performance with the conventional DPD models. Therefore, it is urgent to develop new DPD models with better nonlinear fitting capability.

In recent years, owing to its excellent modeling ability [14], neural network (NN) has drawn the attentions of DPD researchers and is considered as a promising candidate modeling method for DPD. Various neural network based models have been developed for DPD [11], [13], [15]–[20]. Multilayer perceptron (MLP) is a representative model of neural networks. In [15], two complex-valued MLPs were employed to model the amplitude modulation to amplitude modulation (AM-AM) and the amplitude modulation to phase modulation (AM-PM) characteristics of PAs separately. However, it's difficult for these two complex-valued MLPs to converge simultaneously. Besides, complex-valued networks have the complex differentiable restriction on activations and necessitate a cumbersome training algorithm with complex gradient operations [21]. Thus, it may lead to a lengthy training time and increased computation resources [16]. To avoid these problems, a real-valued time-delay neural network (RVTDNN) model was developed in [16]. By splitting the input and output into in-phase and quadrature (I/Q) parts, RVTDNN can model the nonlinear characteristics of the RF PAs with one neural network. With the same input and output configurations as those in RVTDNN, many real-valued neural networks for PA modeling and DPD were proposed in different application scenarios [17]–[20]. To mitigate the PAs' nonlinearities with the I/Q imbalance and crosstalk in multi-input multi-output (MIMO) transmitters, composite DPD neural network for MIMO was studied in [20]. Another classic neural network model, long short-term memory (LSTM) model, is widely used to model long sequence data due to the ability to store long memories [22]. To model wideband PAs with strong memory effects, different LSTM models have been used in DPD [13], [23], [24].

It should be emphasized that all the neural network based models mentioned above split the input and output into in-phase and quadrature parts, which violates the "first-zone constraint" and doesn't match the physical mechanisms of PAs [25]. Recently, a novel vector decomposed time-delay neural network (VDTDNN) was proposed [26]. VDTDNN model conducts nonlinear operations on the magnitudes of

the complex input and recovers the phase information by well-designed linear weighting operations. Compared with the neural networks mentioned before, VDTDNN conforms more with physical mechanism of nonlinear PAs. Consequently, VDTDNN performs better in behavioral modeling of PAs and digital predistorters compared with other MLP based neural networks.

To fully utilize the modeling capability of LSTM models for RF PAs with wideband inputs, this work combines the vector decomposition mechanism with LSTM models and proposes a novel vector decomposed LSTM (VDLSTM) model. VDLSTM model can outperform conventional models and VDTDNN model in linearizing PAs with wideband inputs. Furthermore, by designing a new hidden state of the LSTM cell, a simplified vector decomposed LSTM (SVDLSTM) model is proposed. Compared to VDLSTM model, SVDLSTM model can maintain comparable performance with much lower model complexity.

The rest of this paper is organized as follows. Section II introduces the proposed VDLSTM model. The proposed simplified VDLSTM model is described in Section III. The complexity analysis and comparision are presented in Section IV. In section V and VI, the experimental validation and conclusion are presented respectively.

II. VECTOR DECOMPOSED LONG SHORT-TERM MEMORY MODEL
A. REVIEW OF VECTOR DECOMPOSED NEURAL NETWORK
Generally, the signals fed into RF PAs are real-valued bandpass signals and the nonlinear distortions are mainly caused by the time-varied signal envelopes. To accurately characterize the nonlinearities of RF PAs, the baseband equivalent PA behavioral models or the corresponding DPD models should conduct nonlinear operations on signals' envelopes rather than on the real and imaginary parts separately. Besides, they must satisfy fundamental odd-parity and unitary phase constraints, i.e. "first-zone constraint" imposed by the bandpass nature of RF PAs excited by modulated signals having a bandwidth much lower than the carrier frequency [25].

To meet the above requirements, the recently proposed VDTDNN model conducts nonlinear operations on the magnitudes of the input signals. Then the phase information is recovered by the linear weighting of the outputs of MLP neural network to satisfy "first-zone constraint" [26]. Specifically, VDTDNN introduces a phase recovery block (PRB). The PRB conducts linear weighting operations on each hidden neuron's output. The weights are related with input signals' phases of different memory terms. A sub-network of VDTDNN is illustrated in Fig. 1, where only a hidden neuron with a certain memory term is considered. The input signal at time step n is a complex baseband signal denoted by $\tilde{x}(n)$. Then we define the input signal vector with memory depth M to be $\tilde{\mathbf{x}}_n$, where $\tilde{\mathbf{x}}_n = \left[\tilde{x}(n), \tilde{x}(n-1), \cdots, \tilde{x}(n-M) \right]^T$. The nonlinear operation output $A_g(n)$ of the g-th neuron with
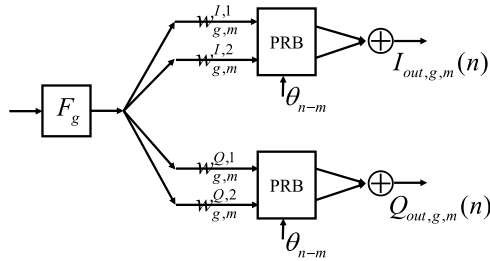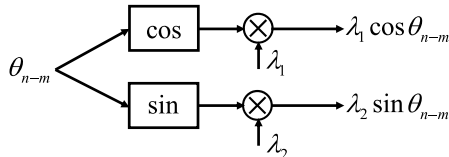
FIGURE 1. The sub-network of VDTDNN.

FIGURE 2. Phase recovery block.

FIGURE 3. The architecture of LSTM cell.

nonlinear function F_g is obtained as follows:

$$A_g(n) = F_g \left[\left| \tilde{\mathbf{x}}_n \right| \right] \qquad (1)$$

where $\left| \tilde{\mathbf{x}}_n \right| = \left[\left| \tilde{x}(n) \right|, \left| \tilde{x}(n-1) \right|, \cdots, \left| \tilde{x}(n-M) \right| \right]^T$ represents the magnitude vector of $\tilde{\mathbf{x}}_n$. Then the output $A_g(n)$ is multiplied by the phase recovery weighting coefficients $w_{g,m}^{I,1}$, $w_{g,m}^{I,2}$, $w_{g,m}^{Q,1}$, $w_{g,m}^{Q,2}$ for the m-th memory term, respectively. After that, in the PRB, these weighting outputs are combined with the complex input signal $\tilde{x}(n-m)$'s phase θ_{n-m} to obtain the in-phase and quadrature parts of the output for the m-th memory term as follows:

$$I_{out,g,m}(n) = w_{g,m}^{I,1} A_g(n) \cos\theta_{n-m} + w_{g,m}^{I,2} A_g(n) \sin\theta_{n-m} \quad (2)$$

$$Q_{out,g,m}(n) = w_{g,m}^{Q,1} A_g(n) \cos\theta_{n-m} + w_{g,m}^{Q,2} A_g(n) \sin\theta_{n-m} \quad (3)$$

Define $\lambda_1 = w_{g,m}^{I,1} A_g(n)$, $\lambda_2 = w_{g,m}^{I,2} A_g(n)$, we take the PRB for Equation (2) in Fig. 2 as an example to illustrate the detailed architecture of PRB. Similarly, in a VDTDNN model consisting of G neurons, the weighted summations for all the neuron outputs after phase recovery with different memory terms are calculated to obtain the in-phase and quadrature parts of final complex output $\tilde{y}(n)$ separately:

$$I_{out}(n) = \sum_{m=0}^{M} \sum_{g=1}^{G} \left(w_{g,m}^{I,1} A_g(n) \cos\theta_{n-m} + w_{g,m}^{I,2} A_g(n) \sin\theta_{n-m} \right)$$

$$= \sum_{m=0}^{M} \left(\mathbf{A}_n^T \mathbf{w}_m^{I,1} \cos\theta_{n-m} + \mathbf{A}_n^T \mathbf{w}_m^{I,2} \sin\theta_{n-m} \right) \quad (4)$$

$$Q_{out}(n) = \sum_{m=0}^{M} \sum_{g=1}^{G} \left(w_{g,m}^{Q,1} A_g(n) \cos\theta_{n-m} + w_{g,m}^{Q,2} A_g(n) \sin\theta_{n-m} \right)$$

$$= \sum_{m=0}^{M} \left(\mathbf{A}_n^T \mathbf{w}_m^{Q,1} \cos\theta_{n-m} + \mathbf{A}_n^T \mathbf{w}_m^{Q,2} \sin\theta_{n-m} \right) \quad (5)$$

where $\mathbf{A}_n = [A_1(n), A_2(n) \cdots, A_G(n)]^T$ represents the neuron output vector, $\mathbf{w}_m^{j,k} = \left[w_{1,m}^{j,k}, w_{2,m}^{j,k}, \cdots, w_{G,m}^{j,k} \right]^T$

represents the phase recovery weighting coefficient vector, $j = I, Q, k = 1, 2$.

As we can see, by decomposing the input signals into magnitudes and phases, VDTDNN model is also based on real numbers and doesn't require complex gradient operations during training. Moreover, compared with conventional real-valued neural networks splitting the input and output into I/Q parts, VDTDNN conforms more with the physical nature of RF PAs due to its vector decomposition mechanism.

B. LONG SHORT-TERM MEMORY MODEL

Recurrent neural network (RNN) is an important branch of neural networks. Due to inherent memory mechanism, RNNs are widely used for modeling sequence data with memory effects. However, when trained with long sequence data, RNNs may encounter the problem of vanishing or exploding gradients. To overcome these problems, a classic variant of RNN, LSTM introduces the cell state to store long memory effects and control gates to control the transmissions of memory effects [22]. The LSTM cell is shown in Fig. 3. Similar to the RNN, at each time step n, the output signal vector of LSTM is the corresponding result of the nonlinear operations on a variable-length sequence $\mathbf{X}_n = [\mathbf{x}_n, \mathbf{x}_{n-1}, \cdots, \mathbf{x}_{n-T}]$, where T represents the sequence length, $\mathbf{x}_n$ represents the input vector at the current time step. The specific forward propagation and recurrent procedures of LSTM model are described as follows [13]:

$$\mathbf{f}_n = \sigma \left(\mathbf{W}_{xf} \mathbf{x}_n + \mathbf{W}_{hf} \mathbf{h}_{n-1} + \mathbf{b}_f \right) \qquad (6)$$

$$\mathbf{i}_n = \sigma \left(\mathbf{W}_{xi} \mathbf{x}_n + \mathbf{W}_{hi} \mathbf{h}_{n-1} + \mathbf{b}_i \right) \qquad (7)$$

$$\mathbf{o}_n = \sigma \left(\mathbf{W}_{xo} \mathbf{x}_n + \mathbf{W}_{ho} \mathbf{h}_{n-1} + \mathbf{b}_o \right) \qquad (8)$$

$$\mathbf{g}_n = \tanh \left(\mathbf{W}_{xg} \mathbf{x}_n + \mathbf{W}_{hg} \mathbf{h}_{n-1} + \mathbf{b}_g \right) \qquad (9)$$

where $\mathbf{f}_n$, $\mathbf{i}_n$, $\mathbf{o}_n$, $\mathbf{g}_n$ represent the forget gate, the input gate, the output gate and nonlinear output, respectively. $\mathbf{W}_{xf}$, $\mathbf{W}_{xi}$, $\mathbf{W}_{xo}$ and $\mathbf{W}_{xg}$ are weight matrices for the current input vector $\mathbf{x}_n$. $\mathbf{W}_{hf}$, $\mathbf{W}_{hi}$, $\mathbf{W}_{ho}$ and $\mathbf{W}_{hg}$ are weight matrices for the previous hidden state $\mathbf{h}_{n-1}$. $\mathbf{b}_f$, $\mathbf{b}_i$, $\mathbf{b}_o$ and $\mathbf{b}_g$ are corresponding bias terms. The current cell state $\mathbf{c}_n$, the hidden state $\mathbf{h}_n$ and

FIGURE 4. The folded architecture of VDLSTM model.

the LSTM cell's final output $\mathbf{A}_n$ can be calculated by:

$$\mathbf{c}_n = \mathbf{f}_n \otimes \mathbf{c}_{n-1} + \mathbf{i}_n \otimes \mathbf{g}_n \qquad (10)$$

$$\mathbf{A}_n = \mathbf{h}_n = \mathbf{o}_n \otimes \tanh(\mathbf{c}_n) \qquad (11)$$

where $\mathbf{c}_{n-1}$ represents the previous cell state, $\otimes$ represents element-wise multiplication.

C. VECTOR DECOMPOSED LSTM MODEL
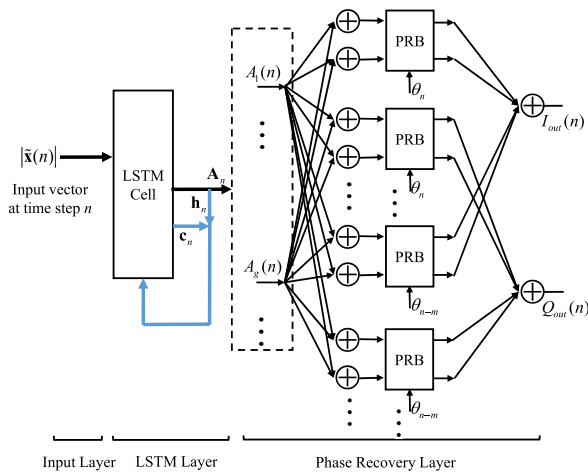
Due to the cell state and the control gates, LSTM model can model long sequence data better compared with MLP model and RNN model. Therefore, LSTM model is a promising candidate that can model the nonlinear behaviors of RF PAs with long memory effects. However, conventional LSTM models for DPD conduct nonlinear operations on in-phase and quadrature parts of complex signals separately, which doesn't comply to the physical mechanisms of nonlinear PAs. Therefore, vector decomposition mechanism is applied in LSTM model and VDLSTM model is proposed.

Similar to VDTDNN model based on MLP neural network, the proposed VDLSTM model conducts nonlinear operations on the magnitudes of the input signals and recovers the phase information by linear weighting of the outputs of the LSTM cell. The folded architecture of VDLSTM model is illustrated in Fig. 4.

It should be noted that there is a difference in the form of the input signals between MLP based VDTDNN model and LSTM based VDLSTM model. The input of VDLSTM model is a sequence. At each time step, the current element of the sequence enters the LSTM cell. Due to the recurrent mechanism, VDLSTM model with a certain sequence length T can model the memory effects to some extent. Furthermore, to better model the memory effects of wideband RF PAs, the input of VDLSTM model at current time step n is defined to be $\left|\tilde{\mathbf{x}}_n\right| = \left[\left|\tilde{x}(n)\right|, \left|\tilde{x}(n-1)\right|, \cdots, \left|\tilde{x}(n-M)\right|\right]^T$, i.e. the magnitude vector with memory depth of M, which is the same as the input of VDTDNN. Naturally, the input at the previous time step is $\left|\tilde{\mathbf{x}}_{n-1}\right| = \left[\left|\tilde{x}(n-1)\right|, \left|\tilde{x}(n-2)\right|, \cdots, \left|\tilde{x}(n-1-M)\right|\right]^T$. Accordingly,

the input at the past t-th time step is $\left|\tilde{\mathbf{x}}_{n-t}\right| = \left[\left|\tilde{x}(n-t)\right|, \left|\tilde{x}(n-t-1)\right|, \cdots, \left|\tilde{x}(n-t-M)\right|\right]^T$ where $0 \le t \le T$. Since the memory terms at the past time steps, quantified as T, influence the current LSTM cell's output by the transmission of the hidden state, these memory terms are called indirect memory terms (IDMT) in this paper. Oppositely, the memory terms at the current time step quantified as M, directly influence the current LSTM cell's output and are called direct memory terms (DMT). The input sequence with IDMT and DMT are expressed as:

$$\left|\tilde{\mathbf{X}}_n\right| = \left[\left|\tilde{\mathbf{x}}_n\right|, \cdots, \left|\tilde{\mathbf{x}}_{n-t}\right|, \cdots, \left|\tilde{\mathbf{x}}_{n-T}\right|\right]^T$$

$$= \begin{bmatrix} |\tilde{x}(n)| & |\tilde{x}(n-1)| & \cdots & |\tilde{x}(n-M)| \\ \vdots & \vdots & \ddots & \vdots \\ |\tilde{x}(n-t)| & |\tilde{x}(n-t-1)| & \cdots & |\tilde{x}(n-t-M)| \\ \vdots & \vdots & \ddots & \vdots \\ |\tilde{x}(n-T)| & |\tilde{x}(n-T-1)| & \cdots & |\tilde{x}(n-T-M)| \end{bmatrix}$$

$$(12)$$

To model the relationship between the input sequence $\left|\tilde{\mathbf{X}}_n\right|$ and the complex output $\tilde{y}(n)$, we apply the sequence-to-one regression mode to the proposed VDLSTM, which is depicted in Fig. 5. Firstly, based on Equation (6)-(11), the corresponding output of the LSTM cell at the last time step n, denoted by $\mathbf{A}_n$ here can be calculated with $\left|\tilde{\mathbf{X}}_n\right|$ iteratively. In VDLSTM model with G hidden neurons, $\mathbf{A}_n = [A_1(n), A_2(n), \cdots, A_G(n)]^T$ has a size of $G \times 1$. Secondly, these hidden neurons' outputs are multiplied by the phase recovery weighting coefficients separately. Thirdly, after weighting operations, the weighting outputs are recovered with phase information in multiple PRBs. Finally, the in-phase part $I_{out}(n)$ and quadrature part $Q_{out}(n)$ of the final complex output $\tilde{y}(n)$, can be obtained as Equation (4) and (5). The phase recovery operations in VDLSTM model are similar to those in VDTDNN model.

III. SIMPLIFIED VECTOR DECOMPOSED LONG SHORT-TERM MEMORY MODEL

In this section, a novel simplified VDLSTM model is proposed. In the LSTM cell of VDLSTM model, the size of the

FIGURE 5. Sequence-to-one regression mode of VDLSTM model.

hidden state is the number of hidden neurons G. Since the number of parameters of the weight matrices for the previous hidden state is proportional to G^2, the total number of parameters will become quite large when there are numerous hidden neurons in VDLSTM model. Therefore, the priority of VDLSTM model's simplification is to reduce the size of the hidden state without modeling performance degradation. Thus, the new hidden state with fewer elements needs to present good representation ability of memory effects comparable to the hidden state $\mathbf{h}_n$ in VDLSTM model. In VDLSTM model, the output $\mathbf{A}_n$ of the LSTM cell that needs to be recovered with phase information is the same as the hidden state $\mathbf{h}_n$. Hence, the intermediate variables that are generated during the phase recovery stage contain the information of the hidden state and some of them could be used as the elements in the new hidden state. Based on the above analysis, the detailed design processes of the new hidden state are described as follows. As we can see, the complex output of VDLSTM model with phase recovery is expressed as:

$$
\begin{aligned}
\tilde{y}(n) &= I_{out}(n) + jQ_{out}(n) \\
&= \sum_{m=0}^{M} \left(\mathbf{A}_n^T \mathbf{w}_m^{I,1} \cos\theta_{n-m} + \mathbf{A}_n^T \mathbf{w}_m^{I,2} \sin\theta_{n-m} \right) \\
&\quad + j \sum_{m=0}^{M} \left(\mathbf{A}_n^T \mathbf{w}_m^{Q,1} \cos\theta_{n-m} + \mathbf{A}_n^T \mathbf{w}_m^{Q,2} \sin\theta_{n-m} \right)
\end{aligned} \tag{13}
$$

We define the in-phase and quadrature parts of the m-th memory term as:

$$
I_{out,m}(n) = \mathbf{A}_n^T \mathbf{w}_m^{I,1} \cos\theta_{n-m} + \mathbf{A}_n^T \mathbf{w}_m^{I,2} \sin\theta_{n-m} \tag{14}
$$

$$
Q_{out,m}(n) = \mathbf{A}_n^T \mathbf{w}_m^{Q,1} \cos\theta_{n-m} + \mathbf{A}_n^T \mathbf{w}_m^{Q,2} \sin\theta_{n-m} \tag{15}
$$

Then the complex output of VDLSTM can be rewritten as:

$$
\begin{aligned}
\tilde{y}(n) &= \sum_{m=0}^{M} \left[I_{out,m}(n) + jQ_{out,m}(n) \right] \\
&= \sum_{m=0}^{M} |\tilde{y}_m(n)| \, e^{j\angle \tilde{y}_m(n)}
\end{aligned} \tag{16}
$$

where $|\tilde{y}_m(n)|$ represents the output magnitude of the m-th memory term, $\angle \tilde{y}_m(n)$ represents the corresponding phase. Apparently, both the output magnitude $|\tilde{y}_m(n)|$ and the corresponding phase $\angle \tilde{y}_m(n)$ contain information of the hidden state. Considering that the nonlinear operations of VDLSTM model should be conducted on magnitudes rather than on in-phase and quadrature components separately, the hidden state transmitted to the next time step should be composed of magnitudes. Therefore, we define a new hidden state composed of output magnitudes of all the memory terms as:

$$
\mathbf{h}_n^{PRB} = \left[A_0^{PRB}(n), A_1^{PRB}(n), \cdots, A_m^{PRB}(n), \cdots, A_M^{PRB}(n) \right]^T
$$

where:

$$
A_m^{PRB}(n) = |\tilde{y}_m(n)| = \sqrt{I_{out,m}(n)^2 + Q_{out,m}(n)^2} \tag{17}
$$

FIGURE 6. Modified LSTM cell combined with phase recovery operations.

FIGURE 7. The folded architecture of SVDLSTM model.

The size of new hidden state is $M+1$. Generally, the number of the hidden neurons G is much larger than the memory depth M. Thus, the size of the new hidden state is greatly reduced compared to that in VDLSTM model. Besides, even though the new hidden state is derived during the phase recovery stage, it doesn't violate the vector decomposition mechanism.

To transmit the new hidden state composed of output magnitudes, the output of original LSTM cell needs to be recovered with phase information and the output magnitudes of different memory terms are calculated in every time steps. Therefore, in the proposed SVDLSTM model, the original LSTM cell architecture is modified by adding phase recovery operations inside the LSTM cell as Fig. 6. With the modified LSTM cell, the folded architecture of SVDLSTM model is illustrated in Fig. 7. As shown in Fig. 7, the modified LSTM layer with phase recovery operations is composed of an LSTM cell, a fully-connected layer and $M+1$ recurrent PRBs for each memory term respectively. The LSTM cell conducts nonlinear operations on magnitudes of the input signal. Then, similar to the phase recovery operations in VDTDNN and VDLSTM models, the fully-connected layer conducts linear weighting operations on the LSTM cell's output $\mathbf{A}_n$. At last, the weighting outputs are recovered with phase information in the novel recurrent PRBs. In recurrent PRBs, not only in-phase and quadrature parts of final complex output are
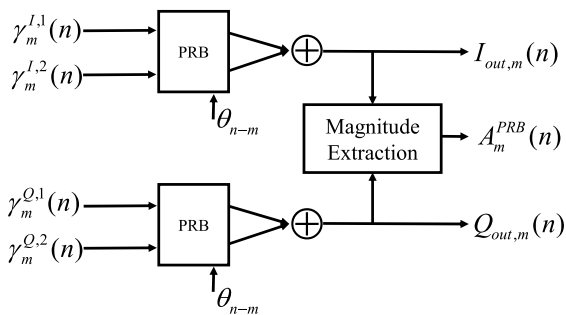
FIGURE 8. The architecture of recurrent PRB.

calculated, but also the new hidden state $\mathbf{h}_n^{PRB}$ is obtained for the input at the next time step. The detailed architecture of a recurrent PRB is shown in Fig. 8. A recurrent PRB for the m-th memory term consists of two PRBs. We define the inputs of the recurrent PRB, i.e. the weighting LSTM cell's outputs to be $\gamma_m^{I,1}(n) = \mathbf{A}_n^T \mathbf{w}_m^{I,1}$, $\gamma_m^{I,2}(n) = \mathbf{A}_n^T \mathbf{w}_m^{I,2}$, $\gamma_m^{Q,1}(n) = \mathbf{A}_n^T \mathbf{w}_m^{Q,1}$, $\gamma_m^{Q,2}(n) = \mathbf{A}_n^T \mathbf{w}_m^{Q,2}$. The first two inputs $\gamma_m^{I,1}(n)$, $\gamma_m^{I,2}(n)$ enter the first PRB, while the last two inputs $\gamma_m^{Q,1}(n)$, $\gamma_m^{Q,2}(n)$ enter the second PRB. Then with the outputs of these two PRBs, the in-phase part $I_{out,m}(n)$ and quadrature part $Q_{out,m}(n)$ for the m-th memory can be calculated as Equation (14) and (15). The output magnitude $A_m^{PRB}(n)$ of the m-th memory term is calculated as (17) with $I_{out,m}(n)$ and $Q_{out,m}(n)$ in the magnitude extraction block. The output magnitude $A_m^{PRB}(n)$ is used as an element in the new hidden state $\mathbf{h}_n^{PRB}$, and the I/Q components $I_{out,m}(n)$ and $Q_{out,m}(n)$ are used to calculate the in-phase and quadrature parts of the final output $\tilde{y}(n)$ respectively. According to the above-mentioned description, the construction of SVDLSTM model is finished. The complete forward propagation and recurrent procedures are presented as follows. First, by replacing the original hidden state with the new hidden state, Equation (6)-(9) can be modified as:

$$\mathbf{f}_n = \sigma\left(\mathbf{W}_{xf}\mathbf{x}_n + \mathbf{W}_{hf}^{PRB}\mathbf{h}_{n-1}^{PRB} + \mathbf{b}_f\right) \tag{18}$$

$$\mathbf{i}_n = \sigma\left(\mathbf{W}_{xi}\mathbf{x}_n + \mathbf{W}_{hi}^{PRB}\mathbf{h}_{n-1}^{PRB} + \mathbf{b}_i\right) \tag{19}$$

$$\mathbf{o}_n = \sigma\left(\mathbf{W}_{xo}\mathbf{x}_n + \mathbf{W}_{ho}^{PRB}\mathbf{h}_{n-1}^{PRB} + \mathbf{b}_o\right) \tag{20}$$

$$\mathbf{g}_n = \tanh\left(\mathbf{W}_{xg}\mathbf{x}_n + \mathbf{W}_{hg}^{PRB}\mathbf{h}_{n-1}^{PRB} + \mathbf{b}_g\right) \tag{21}$$

where $\mathbf{W}_{hf}^{PRB}$, $\mathbf{W}_{hi}^{PRB}$, $\mathbf{W}_{ho}^{PRB}$ and $\mathbf{W}_{hg}^{PRB}$ are new weight matrices for the previous hidden state. Then, the output of the LSTM cell composed of all the hidden neurons' outputs, can be obtained as (10) and (11). After phase recovery and magnitude extraction operations, the current hidden state $\mathbf{h}_n^{PRB}$ that will be transmitted to the next time step and I/Q components $I_{out}(n)$, $Q_{out}(n)$ of final output can be calculated.

IV. COMPLEXITY ANALYSIS AND COMPARISON OF VDLSTM AND SVDLSTM
In this section, the complexity analysis of the proposed VDLSTM model and SVDLSTM model is presented.
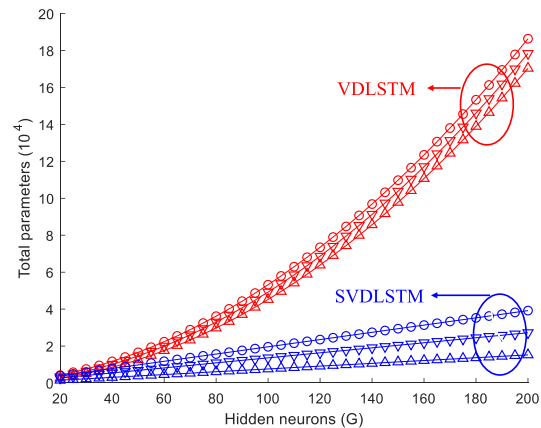

FIGURE 9. Relationship between total parameters and the number of hidden neurons. The upper triangular denotes $M = 5$, the lower triangular denotes $M = 10$, the circle denotes $M = 15$.

Compared to VDLSTM, SVDLSTM has much fewer parameters while maintaining equivalent performance.

We assume that there is only one hidden layer in VDLSTM and SVDLSTM models, and the number of hidden neurons is G. The memory depth of the input vector is M. Since the input sequence length T does not bring differences to the total parameters of VDLSTM and SVDLSTM models, it is not considered here. There are two parts of parameters: (i) The weight matrices for the current input and the previous hidden state in the LSTM cell, together with the corresponding bias terms; (ii) The phase recovery coefficients in the PRBs or the recurrent PRBs. The number of total parameters in VDLSTM model is:

$$\begin{aligned} N_1 &= 4 \times ((M+1) + G + 1) \times G + 4 \times G \times (M+1) \\ &= 4G^2 + 8MG + 12G \end{aligned} \tag{22}$$

While the number of parameters in SVDLSTM model is:

$$\begin{aligned} N_2 &= 4 \times ((M+1) \times 2 + 1) \times G + 4 \times G \times (M+1) \\ &= 12MG + 16G \end{aligned} \tag{23}$$

Then the quantity gap of parameters between VDLSTM model and SVDLSTM model can be obtained as:

$$N_1 - N_2 = 4 \times (G - M - 1) \times G \tag{24}$$

Equation (22) for VDLSTM model includes a term that is proportional to G^2, while Equation (23) for SVDLSTM model only includes terms that are proportional to MG or G separately. Generally, to realize a relatively good modeling performance, the number of hidden neurons G is much larger than the memory depth M especially when modeling PAs with wideband inputs. This fact means that the network complexity of SVDLSTM model is much lower than VDLSTM model. Taking $G = 100$ and $M = 6$ as an example, the total parameters of VDLSTM and SVDLSTM models are 46000 and 8800 respectively. The visualization of the relationship between the total parameters and the number of hidden neurons G with different memory depths M are
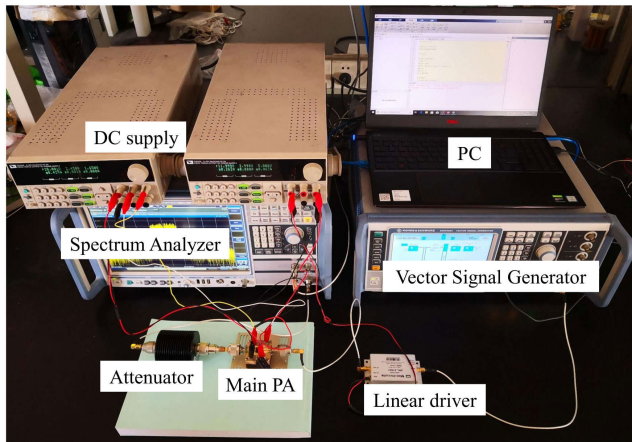
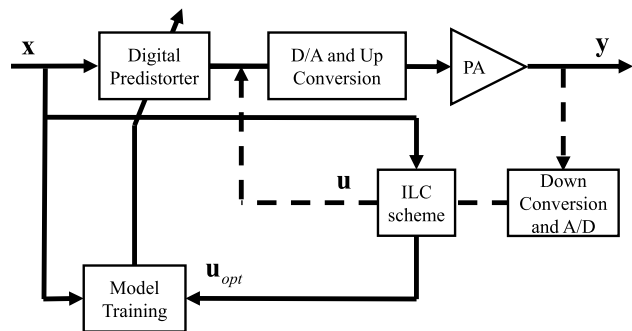**FIGURE 10.** Photograph of the experiment bench.

FIGURE 11. The flow diagram of ILC-based DPD.

illustrated in Fig. 9. Equation (24) shows that as G becomes larger to model more complex nonlinearities, the quantity gap becomes much larger between these two models.

V. EXPERIMENTAL VALIDATION

In this section, the modeling and linearization performance of the conventional models, the proposed VDLSTM and SVDLSTM models are compared. The photograph of the experiment platform is shown in Fig. 10, which consists of a personal computer (PC) with MATLAB and Python softwares, a Vector Signal Generator (SMW200A) from Rohde and Schwarz, a linear driver amplifier, a main PA under test, a 40 dB RF attenuator and a Spectrum Analyzer (FSW43) from Rohde and Schwarz. The main PA is a 2-stage fully integrated Doherty PA working at 2.4 GHz from Ampleon company. The specific transmitting and receiving procedures are explained as follows: First, the baseband signals are generated and up-converted to the carrier frequency of 2.4 GHz by the Vector Signal Generator under the control of the PC; Second, after being amplified by the linear driver amplifier, the up-converted RF signals are sent into the Doherty PA that needs to be linearized. Third, after being attenuated by the attenuator, the PA outputs are down-converted and captured by the Spectrum Analyzer and sent to the PC.

In this article, iterative learning control (ILC) method is used to verify the linearization performance of the proposed models. The main idea of ILC is changing adaptively the input of the unknown system to minimize the error between actual system output and ideal system output [27]. As depicted in Fig. 11, ILC is applied to DPD identification, the unknown system is the PA, the ideal system output is the linear amplification of the input of the digital predistorter **x**, the system input is the predistorted signal **u**. First, based on ILC algorithm, the PA's input signal **u** is changed adaptively to minimize the error between **x** and the actual PA output **y**, which is illustrated by the dashed line in Fig. 11. By using ILC algorithm, the optimal PA input, i.e. the optimal predistorted signal $\mathbf{u}_{opt}$ is obtained. Second, by selecting a proper model, the digital predistorter can be identified based on the input of the digital predistorter and the optimal predistorted signal. Third, the derived digital predistorter gives the new predistorted signal and sends it into the PA to realize the linearization of the cascade of the digital predistorter and the PA.

A. MODELING PERFORMANCE COMPARISON

In this experiment, in order to compare the modeling ability of the proposed VDLSTM model and the conventional models, the widely used GMP and DVR models, and the VDTDNN model are used for comparison. These models are used to model the digital predistorter, i.e. the pre-inverse of the PA that needs to be linearized. We use ILC method mentioned above to obtain the optimal predistorted signal for the identification of the digital predistorter.

The test signal is a 5-carrier 100 MHz OFDM signal with PAPR of 6.6 dB. The main PA is driven under a high compression and the corresponding nonlinear gain is reduced by 2 dB compared to the PA's small signal linear gain. The output power of the main PA is 37.1 dBm. Around 50000 samples were recorded with sampling rate at 368.64 MSPS in each iteration of ILC scheme to obtain the optimal predistorted signal. The test signal and the optimal predistorted signal are then used to train the models. The configurations of defferent models are described as follows: The nonlinear order and memory depth of GMP model here are 13 and 9 respectively, the nonlinear order and memory depth for the lagging and leading cross terms are 9 and 7 respectively, and the orders of lagging and leading cross terms are both 7. DVR model has 13 partitions with memory depth of 13, and it includes linear terms, 1st-order basis, 2nd-order type-1 terms, 2nd-order type-2, 2nd-order type-3 terms and DDR-1 terms introduced in [9]. Both VDTDNN and VDLSTM models here have 1 layer with 100 hidden neurons. The memory depths of the input vectors of VDTDNN and VDLSTM models are both 15, while the input sequence length of VDLSTM model is 7. It should be noted that the the configurations of GMP and DVR models here have presented their best performance, which means that performance of these two models won't be improved with more coefficients. GMP and DVR models

TABLE 1. Modeling performance comparison of different models.

Model Type	GMP	DVR	VDTDNN	VDLSTM
NMSE(dB)	-35.37	-34.21	-33.06	-38.49

TABLE 2. Linearization performance comparison between the state-of-the-art models and VDLSTM model.

Model Type	NMSE (dB)	ACPR(dB) (lower/upper)
without DPD	-20.14	-26.77/-23.67
GMP	-38.57	-46.45/-43.32
DVR	-37.88	-44.32/-43.34
VDTDNN	-37.59	-43.63/-43.00
VDLSTM	**-41.10**	**-48.44/-47.66**

are identified in MATLAB while the trainings of VDTDNN and VDLSTM models are implemented in PyTorch. The maximum iterations of VDTDNN and VDLSTM models are both 600. The modeling performance, denoted by normalized mean square error (NMSE) [27] between the modeling predistorted signal and the optimal predistorted signal of different model is listed in Table 1.

It is clear that compared with the existing models, the novel VDLSTM model can improve NMSE at least 3 dB. Therefore, in wideband scenarios, VDLSTM model is a promising behavioral model for modeling digital predistorters and PAs.

B. LINEARIZATION PERFORMANCE COMPARISON OF CONVENTIONAL MODELS AND VDLSTM MODEL

After the digital predistorter is derived by modeling, the corresponding predistorted signal can be sent into PA to realize the linearization of the PA output signal. In this experiment, the linearization performances of all the models in part A are compared. All the settings in this part are the same as those in part A in this section.

First, the characteristic curves of the AM-AM and AM-PM without DPD and with VDLSTM model are shown in Fig. 12. The detailed NMSE between the linearized PA output and the ideal PA output, and adjacent channel power ratio (ACPR) [27] of the linearized PA output are listed in Table 2. The power spectral density (PSD) comparision is illustrated in Fig. 13. According to the Table 2 and Fig. 13, by utilizing the vector decomposition mechanism, NMSE of VDLSTM model can be improved by at least 2.5 dB compared with that of GMP, DVR and VDTDNN. ACPR of VDLSTM model at the lower sideband can be improved by at least 2 dB, while ACPR at the upper sideband can be improved by at least 4.3 dB.

C. LINEARIZATION PERFORMANCE COMPARISION OF VDLSTM MODEL AND SVDLSTM MODEL

To further validate the linearization performance of the proposed vector decomposition based neural networks for
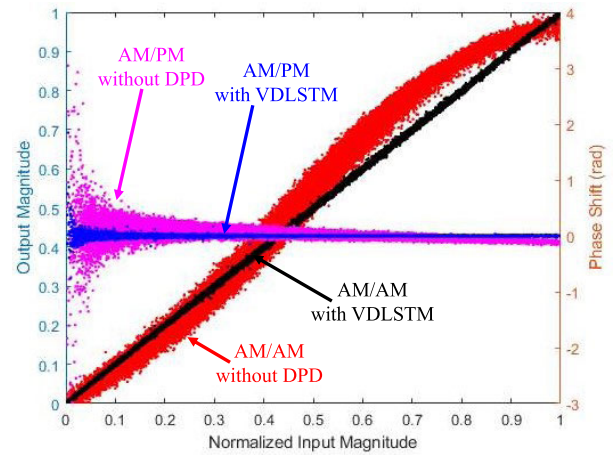
FIGURE 12. AM-AM and AM-PM characteristics with and without DPD for a 100-MHz OFDM signal.
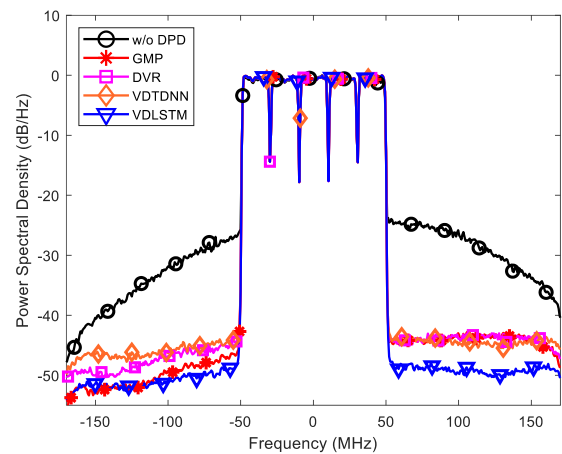
FIGURE 13. Power spectral density comparison between the state-of-the-art models and VDLSTM model.

wideband PAs and to compare the performance between VDLSTM model and SVDLSTM model, a 120 MHz OFDM signal is used as the test signal in this experiment. The settings of VDLSTM model and other experiment settings in this part are the same as those in part A and part B. For the sake of fairness, the settings of SVDLSTM model are the same as those of VDLSTM model.

First, the characteristic curves of the AM-AM and AM-PM without DPD and with SVDLSTM model are shown in Fig. 14. The curves without DPD for a 120 MHz signal are more diffuse compared with those for a 100 MHz signal, which means that the PA has more severe nonlinear distortions with wider signal bandwidth. The NMSEs, ACPRs and total coefficients of different vector decomposition based neural networks are presented in Table 3. The power spectral density comparison is shown in Fig. 15. It can be seen from Table 3 and Fig. 15, both VDLSTM model and SVDLSTM model can realize excellent DPD performance for a 120 MHz signal. Besides, SVDLSTM model can realize comparable

FIGURE 14. AM-AM and AM-PM characteristics with and without DPD for a 120-MHz OFDM signal.

TABLE 3. Linearization performance comparison between VDLSTM model and SVDLSTM model.

Model Type	NMSE (dB)	ACPR(dB) (lower/upper)	Total parameters
without DPD	-20.34	-27.69/-25.59	N/A
VDLSTM	-38.97	-45.39/-45.54	53200
SVDLSTM	-38.75	-44.66/-45.04	19600

FIGURE 15. Power spectral density comparison between VDLSTM model and SVDLSTM model.

linearization performance with 63 percent reduction of total parameters in this experiment.

VI. CONCLUSION

In this research, the vector decomposition mechanism is first applied to LSTM neural network and two novel vector decomposed LSTM neural networks are proposed for behavioral modeling and digital predistortion. The proposed VDLSTM model only conducts nonlinear operations on the magnitudes of the input signals and restores the phase information by linear weighting operations. Compared with the state-of-the-art models such as GMP, DVR and VDTDNN,

the proposed VDLSTM model can achieve better modeling and DPD performance in wideband scenarios. Besides, the proposed SVDLSTM model modifies VDLSTM model by designing a new hidden state that has fewer elements and is still compatible with the vector decomposition mechanism. SVDLSTM model can realize comparable linearization performance compared to VDLSTM model but with significantly less coefficients.

REFERENCES

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

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

[3] 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.

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

[5] 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.

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

[7] 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.

[8] 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. Wireless Compon. Lett.*, vol. 19, no. 8, pp. 530–532, Aug. 2009.

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

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

[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] P. L. Gilabert, G. Montoro, D. Lopez, N. Bartzoudis, E. Bertran, M. Payaro, and A. Hourtane, "Order reduction of wideband digital predistorters using principal component analysis," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2013, pp. 1–4.

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

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

[15] A. Ahmed, E. 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.

[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] 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.

[19] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan. 2019.

[20] 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.*, pp. 1–10, Nov. 2018.

[21] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. Pal, "Deep complex networks," in *Proc. Int. Conf. Learn. Representations*, 2018, pp. 1–19.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[23] P. Chen, S. Alsahali, A. Alt, J. Lees, and P. J. Tasker, "Behavioral modeling of GaN power amplifiers using long short-term memory networks," in *Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits (INMMIC)*, Jul. 2018, pp. 1–3.

[24] H. Yu, G. Xu, T. Liu, J. Huang, and X. Zhang, "A memory term reduction approach for digital pre-distortion using the attention mechanism," *IEEE Access*, vol. 7, pp. 38185–38194, 2019.

[25] 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.

[26] Y. Zhang, Y. Li, F. Liu, and A. Zhu, "Vector decomposition based time-delay neural network behavioral model for digital predistortion of RF power amplifiers," *IEEE Access*, vol. 7, pp. 91559–91568, 2019.

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

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 September 2018 to July 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.
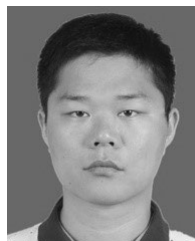
GANG LI received the B.E. degree in electronic engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2013, where he is currently pursuing the Ph.D. degree in electromagnetic field and microwave technology. His current research interests include digital predistortion and nonlinear modeling of transmitters.

FALIN LIU was born in Xingtai, China, in 1963. He received the B.E. degree from Tsinghua University, Beijing, China, in 1985, and the M.E. and Ph.D. degrees in electronic engineering from the University of Science and Technology of China (USTC), Hefei, China, in 1988 and 2004, respectively.

From 1997 to 1998, he was a Visiting Scholar with Tohoku University, Sendai, Japan. Since 1988, he has been with the Department of Electronic Engineering and Information Science, USTC, where he is currently a Full Professor. He has authored over 90 articles in refereed journals and international conferences. His current research interests include millimeter-wave transceivers, computational electromagnetics, microwave devices and communications, and radar imaging.

Dr. Liu is a Senior Member of the Chinese Institute of Electronics, Beijing. He was a recipient of the Second Prize of the National Science and Technology Progress Award and the First Prize of the CAS Science and Technology Progress Award. He is an Associate Editor-in-Chief of the *Journal of Microwaves* (in Chinese).

• • •

HONGMIN LI received the B.E. degree in electronic information engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2016, where he is currently pursuing the Ph.D. degree in electromagnetic field and microwave technology. His research interests include digital predistortion of power amplifiers, nonlinear system identification, and machine learning.