IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 17, 2020, accepted April 4, 2020, date of publication April 9, 2020, date of current version April 22, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.2986816*

# Instant Gated Recurrent Neural Network Behavioral Model for Digital Predistortion of RF Power Amplifiers

**GANG LI, YIKANG ZHANG, HONGMIN LI, WEN QIAO, AND FALIN LIU**
Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China
Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei 230027, China

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

This work was supported by the National Natural Science Foundation of China under Grant 61471333.

**ABSTRACT** This article presents two novel neural network models based on recurrent neural network (RNN) for radio frequency power amplifiers (RF PAs): instant gated recurrent neural network (IGRNN) model and instant gated implict recurrent neural network (IGIRNN) model. In IGRNN model, two state control units are introduced to ensure the linear transmission of hidden state and solve the problem of vanishing gradients of RNN model. In contrast with conventional RNN model, IGRNN can better describe the long-term memory effect of power amplifier, more in line with the physical distortion characteristics of power amplifier. Furthermore the instantaneous gates are used to express the input information implicitly to reduce the redundancy of the input information, and a simpler IGIRNN model is proposed. The complexity analysis indicates that the proposed models have significantly lower complexity than other RNN-based variant structures. A wideband Doherty RF PA excited by 100MHz and 120MHz OFDM signals was employed to evaluate the performance. Extensive experimental results reveal that the proposed IGRNN and IGIRNN models can achieve better linearization performance compared with RNN model and traditional GMP model, and have comparable performance with lower computational complexity compared with the state-of-the-art RNN-based variant models, such as gated recurrent unit (GRU) model.

**INDEX TERMS** Nonlinear RF PA, digital predistortion, recurrent neural network, instant gated, behavioral modeling.

## I. INTRODUCTION

With the arrival of the fifth-generation (5G) wireless communication system, the system capacity and communication rate are expected to increase significantly [1], [2]. And in order to meet the requirements of high capacity and high rate of the system, the signal will have wider bandwidth and more complex modulation, which will lead to a higher peak to average ratio (PAPR) seriously affecting the linearity and efficiency of radio frequency power amplifiers (RF PAs) with the inherent nonlinear characteristics. There are many approaches aiming to linearize the RF PA and keep a high efficency at the same time, including feedback linearization [3], feedforward linearization [4], analog predistortion and digital predistortion (DPD) [5]. Among the linearization techniques, DPD is generally believed to be the most powerful linearization technology for its flexibility and high performance. At present, scholars and engineers are very interested in the accurate modeling and linearization of wideband RF PA.

Lots of DPD models have been proposed to compensate the nonlinearities of RF PAs [6]–[10]. Volterra-based models such as memory polynomial (MP) [6], generalized memory polynomial (GMP) [7] and dynamic deviation reduction Volterra (DDR) model [8] are the most widely used. In virtue of the correlation of basis function of Volterra-based models, based on canonical piecewise-linear (CPWL) functions, decomposed vector rotation (DVR) is proposed to model the PAs with strong nonlinearity such as envelope tracking (ET) PAs [9]. Although these models have a good performance in narrowband, their linearization performance deteriorates to some extent with the increase of signal bandwidth. In addition, to meet the needs of the industry, many advanced

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

PA architectures have been proposed, such as out-phasing PA, distributed PA and multistage Doherty PA. Some of them may be quite different from the traditional PA in terms of behavior characteristics, so more powerful models are needed to describe different PAs behavior characteristics [10].

Because neural network has good nonlinear fitting ability [11], it is considered to be a potential DPD method. Up to now, some DPD models based on neural network have been developed [12]–[17]. A real-valued time-delay neural network (RVTDNN) model based on multilayered perceptron (MLP) structure is proposed [12], in which input and output are divided into in-phase and quadrature (I/Q) parts to avoid complex gradient operation. Another vector decomposed time-delay neural network (VDTDNN) model based on MLP network is proposed [13], which is characterized by nonlinear operation only on the amplitude of the input signal, and by linear weighting to recover the phase information. However, when it comes to the wideband communication scenarios, the memory effect and nonlinearities of RF PAs will be much more severe and complex, which leads to the relatively inferior linearization performance with MLP architecture. It should be noted that the memory effect of power amplifier will increase significantly with the increase of bandwidth [15]. Recently, the recurrent neural network (RNN) is used to model the behavior of the PA [14], because of its ability of modeling memory effect. However, there is a problem of vanishing gradients with the RNN, implying that the model can not express long-term memory effect well [18], [19]. Therefore, even though RNN model has stronger modeling ability than MLP model in theory, RNN model usually has weaker performance in modeling the behavioral characteristics of wideband RF PAs in practice. In order to solve the problem of gradient vanishing, a few RNN based variant models have been proposed, such as long short-term memory (LSTM) network [20], [21], gated recurrent unit (GRU) network [22], etc. These RNN-based variant models have a fairly good performance, especially in wideband digital predistortion [15]–[17].

Different neural network topologies are developed to realize DPD. The structure of MLP based models are simple, but the performance in wideband is limited. Many RNN-based variant models have good expression ability in wideband modeling, however, with quite high complexity. Therefore, in this paper, we propose two new models based on RNN model: instant gated recurrent neural network (IGRNN) and instant gated implict recurrent neural network (IGIRNN). Compared with RNN model, the two proposed models can better describe the long-term memory effect, more in line with the physical characteristics of power amplifier. At the same time the structure is simpler and the model parameters are less compared with other RNN-based variant structures, such as GRU model. Furthermore, both theoretical analysis and experimental results reveal that the proposed models have better performance than RNN model and traditional GMP model. Compared with the GRU model, they have comparable performance, but the model parameters and computational complexity are significantly reduced.

The rest of the paper is organized as follows. Section II reviews the RNN network. The proposed IGRNN cell and IGIRNN cell are described in Section III. The behavioral models based on IGRNN and IGIRNN are introduced in Section IV. A computational complexity comparison is analyzed in Section V. The experimental validation and a brief conclusion are presented in Section VI and Section VII, respectively.

## II. REVIEW OF RECURRENT NEURAL NETWORK

RNN is a typical regression network dealing with nonlinear problems. Because of its inherent memory mechanism, it has been successfully applied in many fields. For the sake of illustration, a classical RNN network with one RNN layer and followed by a fully-connected layer is considered here, where detailed architecture of the $i$-th hidden neuron in the RNN layer is pictured in Fig. 1. The structure of the network is similar to the standard multilayer perceptron (MLP), the difference is that RNN allows connection between hidden units related to time delay. Through this connection, the model can retain information about the past, so as to discover the time correlation between events far away from the data. The $i$-th hidden neuron of recurrent neural network, with input vector $\mathbf{x}_t$ and state $\mathbf{h}_{t-1}^i$ for time step $t$, is given by:

$$\mathbf{u}_t^i = \mathbf{h}_t^i = \tanh(\mathbf{W}^i\mathbf{x}_t + \mathbf{U}^i\mathbf{h}_{t-1}^i + \mathbf{b}^i) \tag{1}$$

where the superscript indicates the index of the $i$-th hidden neuron and the subscript $t$ represents the index of the time step. $\mathbf{W}^i$ is the weight matrix for the current input vector $\mathbf{x}_t$ and $\mathbf{U}^i$ is the weight matrix for the previous hidden state $\mathbf{h}_{t-1}^i$. $\mathbf{b}^i$ is the bias terms. $\mathbf{u}_t^i$ is the output vector of the $i$-th hidden neuron for time step $t$.

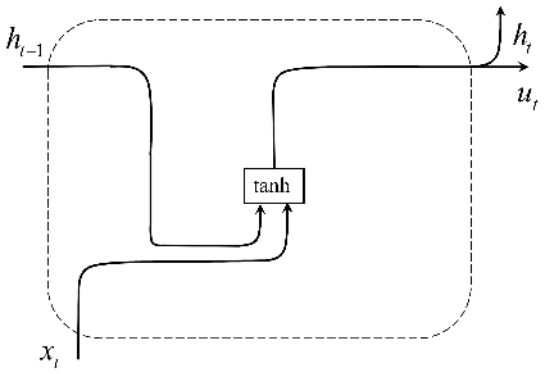



**FIGURE 1.** The architecture of RNN cell.

When the output of RNN layer is obtained, on the one hand, $(\mathbf{u}_t^1, \cdots, \mathbf{u}_t^i, \cdots, \mathbf{u}_t^{N1})$ are transmitted to the next time step hidden state, on the other hand, $(\mathbf{u}_t^1, \cdots, \mathbf{u}_t^i, \cdots, \mathbf{u}_t^{N1})$ pass the next fully-connected layer and it yields the final output

$\mathbf{y}_t = (\mathbf{y}_t^1, \cdots, \mathbf{y}_t^k, \cdots, \mathbf{y}_t^{N2})$. $N1$ and $N2$ are the number of neurons of the RNN layer and the output layer, respectively.The back-propagation through time (BPTT) algorithm is usually adopted to train the model parameters [23]. Output unit $k$'s target at time $t$ is denoted by $\mathbf{d}_t^k$, and the estimated output is $\mathbf{y}_t^k$. At the time step $t$, the cost function can be defined as

$$E_t = \frac{1}{2}(\mathbf{d}_t^k - \mathbf{y}_t^k)^2 \tag{2}$$

The error of the whole network can be defined as

$$E = \sum_{1 \le t \le T} E_t \tag{3}$$

In the case of back propagation, the gradient loss of a time step $t$ is determined by the corresponding gradient loss of the output of the current time step $t$ and the gradient loss of the next time step $t+1$. Thus, the partial derivative of $E$ with respect to $\mathbf{h}_t^j$

$$\frac{\partial E}{\partial \mathbf{h}_t^j} = \sum_k^{N2} \frac{\partial E}{\partial \mathbf{y}_t^k} \frac{\partial \mathbf{y}_t^k}{\partial \mathbf{h}_t^j} + \sum_s^{N1} \frac{\partial E}{\partial \mathbf{h}_{t+1}^s} \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_t^j} \tag{4}$$

$$\frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_t^j} = \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{net}_{t+1}^s} \frac{\partial \mathbf{net}_{t+1}^s}{\partial \mathbf{h}_t^j} \tag{5}$$

where

$$\mathbf{net}_t^s = \mathbf{W}_h^s \mathbf{x}_t + \mathbf{U}_h^s \mathbf{h}_{t-1}^s + \mathbf{b}_h^s \tag{6}$$

With the above recurrence formula, we can get the gradient of the $\tau$-th time step, which can be expressed as

$$\frac{\partial E}{\partial \mathbf{h}_\tau^j} = \sum_k^{N2} \frac{\partial E}{\partial \mathbf{y}_\tau^k} \frac{\partial \mathbf{y}_\tau^k}{\partial \mathbf{h}_\tau^j} + \sum_s^{N1} \frac{\partial E}{\partial \mathbf{h}_{t+1}^s} \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_\tau^j} \tag{7}$$

$$\frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_\tau^j} = \prod_{t \ge i > \tau} \frac{\partial \mathbf{h}_{i+1}^s}{\partial \mathbf{h}_i^j} = \prod_{t \ge i > \tau} \mathbf{U}^T diag(\tanh'(\mathbf{h}_i^j)) \tag{8}$$

The partial derivative of *tanh* is bounded, therefore, when $\tau \ll t+1$, $\left\| \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_\tau^j} \right\|$ will go to zero or infinity [18].If $\left\| \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_\tau^j} \right\| \to 0$, the gradient will vanish, and it means that the long-term memory will not influence the update of current state parameters. Therefore the required parameters cannot be updated accurately. If $\left\| \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_\tau^j} \right\| \to \infty$, this will lead to the problem of gradient explosion. However, gradient clipping can be used to solve the problem by transforming the gradient value into an acceptable range. Therefore when training long-term sequences, RNN networks may encounter the problem of vanishing gradients, which will lead to the difficulty to express long-term memory effects.

## III. INSTANT GATED NOVEL NEURAL NETWORKS

In the field of digital predistortion and RF PAs behavioral modeling, the application of neural network so far is mostly based on MLP network. By reviewing the existing neural network models, RNN model is a classical sequence regression model with simple structure. However, the classical RNN not only suffers from gradient vanishing and explosion problem when coefficients updating, but also is not strictly in accordance with the nonlinear characteristics of RF PAs. Consequently, its DPD linearization performance is relatively limited. Particularly, although some RNN-based variant models, such as GRU, have been designed to avoid gradient vanishing and explosion, these models are all designed for specific application scenarios, such as natural language processing, while still not conform with the RF PAs' nonlinear physical mechanisms. Considering the power consumption and implementation cost, the direct application of GRU model into DPD will lead to very high computation complexity as well as long training time. Therefore, a high performance and low complexity RNN-based model conforming with the nonlinear characteristics of RF PAs should be proposed.

### A. INSTANT GATED RECURRENT NEURAL NETWORK

The static nonlinearity of power amplifier is mainly controlled and influenced by the current input signal. Moreover, with the dynamic self-heating and self-biasing effects, RF PAs shows very strong and complex long-term memory effect under the excitation of high-order modulated broadband signals [24]. Accordingly, the newly-designed DPD models need to better satisfy the nonlinear characteristics of RF PAs, so that both the complex long-term memory effect and the static characteristics under the current input could be precisely reconstructed. For clarity, it is assumed that the current state is $\tilde{\mathbf{h}}_t$, the memory state is $\mathbf{h}_{t-1}$ and the hidden layer state is $\mathbf{h}_t$. Considering that the behavior characteristic of power amplifier is the effect of superposition of memory effect and static working state, the relationship between the final hidden layer state input and these two states can be expressed as follows

$$\mathbf{h}_t = \underset{\uparrow}{\mathbf{W}_1 \otimes \mathbf{h}_{t-1}} + \underset{\uparrow}{\mathbf{W}_2 \otimes \tilde{\mathbf{h}}_t} \tag{9}$$
$$\textit{memory state} \qquad \textit{static state}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are weight matrices for the previous hidden state and the current state, respectively.

In order to adaptively reconstruct the ratio of the static nonlinear effect to dynamic memory effect accurately, a novel state control unit $\mathbf{z}_t$ is introduce here, of which the mathematical expression is shown as:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{b}_z) \tag{10}$$

where $\mathbf{W}_z$ is the weight matrix for the current input $\mathbf{x}_t$. $\mathbf{b}_z$ is the corresponding bias term.

Since the static characteristics is mainly caused by the RF PA excited by the current input signals, combining the state control unit $\mathbf{z}_t$, the static state at instant $t$ could be characterized as followings:

$$\mathbf{W}_2 \otimes \tilde{\mathbf{h}}_t = \mathbf{z}_t \otimes \tilde{\mathbf{h}}_t \tag{11}$$

To get the optimal $\mathbf{h}_t$, the exponential weighted average method in the optimization theory is introduced to adaptively control the ratio of the amount of static nonlinearity to that of dynamic memory nonlinearity. The exponential weighted average method can ensure that $\mathbf{h}_t$ is the optimal value between $\mathbf{h}_{t-1}$ and $\tilde{\mathbf{h}}_t$. Correspondingly, the memory state at instant $t$ could be characterized as followings:

$$\mathbf{W}_1 \otimes \mathbf{h}_{t-1} = (1 - \mathbf{z}_t) \otimes \mathbf{h}_{t-1} \tag{12}$$

Based on Equations (9)-(12), Equation (9) is further rewritten as

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \otimes \mathbf{h}_{t-1} + \mathbf{z}_t \otimes \tilde{\mathbf{h}}_t \tag{13}$$

where $\otimes$ represents element-wise multiplication.

Since the designed network is based on RNN model, naturally, the hidden state output of RNN cell can be assumed as the current state of the cell. In that way, $\tilde{\mathbf{h}}_t$ can be written as

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h \mathbf{h}_{t-1} + \mathbf{b}_h) \tag{14}$$

To make the influence of the previous hidden state $\mathbf{h}_{t-1}$ on update of the current cell state $\tilde{\mathbf{h}}_t$ controllable, a memory state control unit $\mathbf{r}_t$ is also introduced here. Equation (14) is further rewritten as

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h(\mathbf{r}_t \otimes \mathbf{h}_{t-1}) + \mathbf{b}_h) \tag{15}$$

where

$$\mathbf{r}_t = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{b}_r) \tag{16}$$

where $\mathbf{W}_r$ is the weight matrix for the current input $\mathbf{x}_t$. $\mathbf{b}_r$ is the corresponding bias term.

Combined with the analysis above, the forward propagation equation of the $i$-th hidden neuron of the new network structure can be defined as the following form,

$$\mathbf{z}_t^i = \sigma(\mathbf{W}_z^i \mathbf{x}_t + \mathbf{b}_z^i) \tag{17}$$

$$\mathbf{r}_t^i = \sigma(\mathbf{W}_r^i \mathbf{x}_t + \mathbf{b}_r^i) \tag{18}$$

$$\tilde{\mathbf{h}}_t^i = \tanh(\mathbf{W}_h^i \mathbf{x}_t + \mathbf{U}_h^i(\mathbf{r}_t^i \otimes \mathbf{h}_{t-1}^i) + \mathbf{b}_h^i) \tag{19}$$

$$\mathbf{u}_t^i = \mathbf{h}_t^i = (1 - \mathbf{z}_t^i) \otimes \mathbf{h}_{t-1}^i + \mathbf{z}_t^i \otimes \tilde{\mathbf{h}}_t^i \tag{20}$$

where the superscript indicates the index of the $i$-th hidden neuron and the subscript $t$ represents the index of the time step. $\mathbf{W}_z^i$, $\mathbf{W}_r^i$,$\mathbf{W}_h^i$and $\mathbf{U}_h^i$ are corresponding weight matrices. $\mathbf{b}_z^i$, $\mathbf{b}_r^i$and $\mathbf{b}_h^i$ are corresponding bias terms. $\mathbf{u}_t^i$ is output vector of the $i$-th hidden neuron for time step $t$.

Based on Equations (17)-(20), the architecture of the new network model can be illustrated in Fig. 2. In the new structure, due to the introduction of two state control units, the state transfer in the hidden layer has a linear relationship, and the new model avoids the problem of RNN vanishing gradient. We train the new model parameters through back propagation

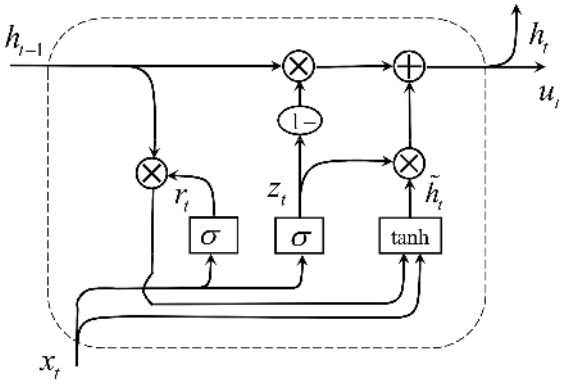



**FIGURE 2.** The architecture of IGRNN cell.

algorithm. Similar to RNN model, we can get the gradient of the hidden layer output,

$$\frac{\partial E}{\partial \mathbf{h}_t^j} = \sum_k^{N2} \frac{\partial E}{\partial \mathbf{y}_t^k} \frac{\partial \mathbf{y}_t^k}{\partial \mathbf{h}_t^j} + \sum_s^{N1} \frac{\partial E}{\partial \mathbf{h}_{t+1}^s} \frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_t^j} \tag{21}$$

$$\frac{\partial \mathbf{h}_{t+1}^s}{\partial \mathbf{h}_t^j} = \frac{\partial \mathbf{h}_{t+1}^s}{\partial \tilde{\mathbf{h}}_{t+1}^s} \frac{\partial \tilde{\mathbf{h}}_{t+1}^s}{\partial \mathbf{net}_{t+1}^s} \frac{\partial \mathbf{net}_{t+1}^s}{\partial \mathbf{h}_t^j} + (1 - \mathbf{z}_{t+1}^j)|_{(s==j)} \tag{22}$$

where

$$\mathbf{net}_t^s = \mathbf{W}_h \mathbf{x}_t + \mathbf{U}_h(\mathbf{r}_t \otimes \mathbf{h}_{t-1}) + \mathbf{b}_h \tag{23}$$

Equation (21) can be rewritten as

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{h}_t^j} = & \sum_k^{N2} \frac{\partial E}{\partial \mathbf{y}_t^k} \frac{\partial \mathbf{y}_t^k}{\partial \mathbf{h}_t^j} \\ & + \sum_s^{N1} \frac{\partial E}{\partial \mathbf{h}_{t+1}^s} \frac{\partial \mathbf{h}_{t+1}^s}{\partial \tilde{\mathbf{h}}_{t+1}^s} \frac{\partial \tilde{\mathbf{h}}_{t+1}^s}{\partial \mathbf{net}_{t+1}^{\tilde{h}s}} \frac{\partial \mathbf{net}_{t+1}^{\tilde{h}s}}{\partial \mathbf{h}_t^j} \\ & + \frac{\partial E}{\partial \mathbf{h}_{t+1}^j}(1 - \mathbf{z}_{t+1}^j) \end{aligned} \tag{24}$$

By comparing Equation (4) with Equation (24), it can be found that Equation (24) adds linear transfer gradient as its last term. If the long-term memory has effect on the current state, the linear return term can ensure that the gradient does not vanish. Therefore, compared with RNN network, the new network solves the problem of vanishing gradients. For Equation (24), the second term on the right hand side can reflect the short-term memory effect, and the third term on the right hand side can reflect the long-term memory effect. Therefore, the new model is more consistent with the behavioral characteristics of RF PA and more powerful to reflect the long-term memory effect of RF PA compared with RNN network.

The new model introduces two state control units, which control the update of the current state and memory state to the hidden layer state through the current input. Contrary to that of GRU, the newly designed model is mainly based on the nonlinear physical characteristics of RF PAs, so that the

new network structure is more in line with the characteristics of power amplifier, making it has excellent expression ability and still with lower computational complexity. In the new structure, the two state control units introduced here only accept that of the current input information $\mathbf{x}_t$. Therefore, these two state control units are called instant gated units. The new network structure is called instant gated recurrent neural network (IGRNN).

### B. INSTANT GATED IMPLICT RECURRENT NEURAL NETWORK

To meet the low cost and low power requirements of digital predistortion, we modify the IGRNN model to make it simpler. From Equation (19), we can find that the update of current state $\tilde{\mathbf{h}}_t^i$ is through the input information $\mathbf{x}_t$, the instant gated $\mathbf{r}_t^i$ and the previous hidden stated $\mathbf{h}_{t-1}^i$, and the update of hidden state $\mathbf{h}_t^i$ is through the instant gated $\mathbf{z}_t^i$ and the current stated $\tilde{\mathbf{h}}_t$. It is observed that two instant gated $\mathbf{z}_t^i$ and $\mathbf{r}_t^i$ already contain the information of current input $\mathbf{x}_t$, where $\mathbf{r}_t^i$ is used for update $\tilde{\mathbf{h}}_t^i$ directly and $\mathbf{z}_t^i$ is used for update $\mathbf{h}_t^i$. As a result, $\tilde{\mathbf{h}}_t^i$ and $\mathbf{h}_t^i$ can indirectly get the input information $\mathbf{x}_t$through $\mathbf{r}_t^i$ and $\mathbf{z}_t^i$. Hence, the current input information can be obtained indirectly by the two gate control units when updating state information. Considering this observation, the current input information $\mathbf{x}_t^i$ in Equation (19) can be removed. Since the two instant gated units are used to express the input information implicitly, this simplified structure is called instant gated implict recurrent neural network (IGIRNN) model. Based on the analysis above, the Equation(19) can be rewritten as

$$\tilde{\mathbf{h}}_t^i = \tanh(\mathbf{U}_h^i(\mathbf{r}_t^i \otimes \mathbf{h}_{t-1}^i) + \mathbf{b}_z^i) \tag{25}$$

The architecture of IGIRNN is illustrated in Fig. 3. Here, the two instant gated units of IGIRNN can be updated by Equation (17) and (18). And the Equation (20) and (25) can be used for the updated of the hidden state $\mathbf{h}_t^i$ and the current cell state $\tilde{\mathbf{h}}_t^i$, respectively.

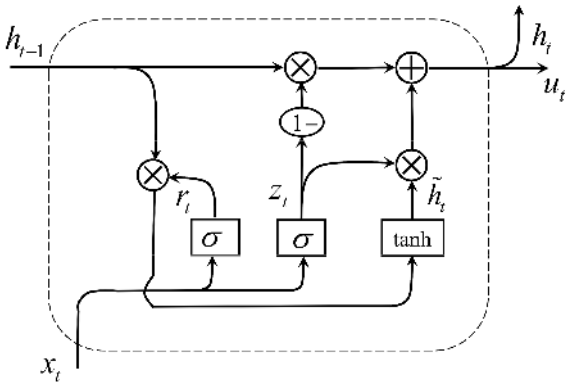



**FIGURE 3.** The architecture of IGIRNN cell.

The IGIRNN still retains the characteristics of linear transmission of hidden layer state, and reduces the redundancy of the input information. Therefore, it can still reflect the behavioral characteristics of RF PA and the influence of long-term memory effect.

## IV. BEHAVIORAL MODELING BASED ON IGRNN OR IGIRNN

When we finished the network cell of IGRNN and IGIRNN construction, the network cell can be expanded to a large one with G hidden neurons in the hidden layer, as is depicted in Fig. 4. The proposed architecture used for behavioral modeling has four layers, namely an input layer, an recurrent neural layer, a fully connected layer and an output layer, as illustrated in Fig. 4.

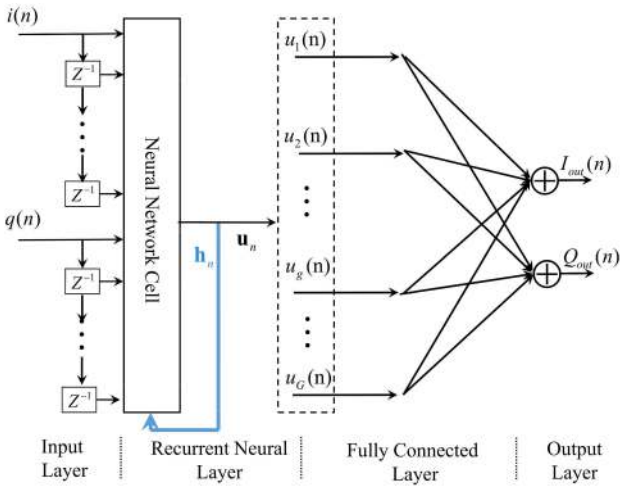



**FIGURE 4.** The folded architecture of behavioral model.

It should be noted that the form of the input signals of both IGRNN and IGIRNN are different from models based on MLP. The input of the proposed models are a sequence. At each time step, the current element of the sequence enters the neural network cell. Due to the recurrent mechanism, the model with a certain sequence length $T$ can model the memory effects to some degree. Further more, to better model the memory effects of wideband RF PAs, the input of the model at current time step $n$ is defined as $\mathbf{x}_n = [x(n), x(n-1), \cdots, x(n-M)]^T$, the input vector with memory depth of $M$, which is the same as the input sequence of the real-valued models based on MLP. Naturally, the input at the previous time step is $\mathbf{x}_{n-1} = [x(n-1), x(n-2), \cdots, x(n-1-M)]^T$, and so forth. Since the memory terms at the past time steps, quantified as $T$, influence the current state by being processed and transmitting the hidden state, these memory terms are called indirect memory terms (IDMT) in this article. Oppositely, the memory terms at the current time steps quantified as M, directly influence the current state, and are called direct memory terms (DMT). Because $\mathbf{x}_n$ is a complex baseband signal, as follows

$$\mathbf{x}_n = \mathbf{i}_n + j\mathbf{q}_n \tag{26}$$

therefore the input sequence $\mathbf{I}_n$and $\mathbf{Q}_n$ with IDMT and DMT are expressed as:

$$\mathbf{I}_n = [\mathbf{i}_n, \mathbf{i}_{n-1}, \cdots, \mathbf{i}_{n-T}] = \begin{bmatrix} i(n) & i(n-1) & \cdots & i(n-M) \\ \vdots & \vdots & \vdots & \vdots \\ i(n-t) & i(n-t-1) & \cdots & i(n-t-M \\ \vdots & \vdots & \vdots & \vdots \\ i(n-T) & i(n-T-1) & \cdots & i(n-T-M) \end{bmatrix} \tag{27}$$

and

$$\mathbf{Q}_n = [\mathbf{q}_n, \mathbf{q}_{n-1}, \cdots, \mathbf{q}_{n-T}] = \begin{bmatrix} q(n) & q(n-1) & \cdots & q(n-M) \\ \vdots & \vdots & \vdots & \vdots \\ q(n-t) & q(n-t-1) & \cdots & q(n-t-M \\ \vdots & \vdots & \vdots & \vdots \\ q(n-T) & q(n-T-1) & \cdots & q(n-T-M) \end{bmatrix} \tag{28}$$

To model the relationship between the input $(\mathbf{I}_n, \mathbf{Q}_n)$ and the output $(I_{out}(n), Q_{out}(n))$, we apply the sequence-to-one regression mode to the proposed models, as shown in Fig. 5. In this mode, based on forward propagation equation of IGRNN or IGIRNN, the corresponding output, denoted by $u_n$ here can be calculated with the input $\mathbf{I}_n$ and $\mathbf{Q}_n$. In the model with $G$ hidden neurons, $\mathbf{u}_n = [u_1(1), y_2(2), \cdots, u_G(n)]$ is the cell's output at the last time step and has a size of $G \times 1$. Then, the in-phase part $\mathbf{I}_{out}(n)$ and quadrature part $\mathbf{Q}_{out}(n)$ of the corresponding cells output, can be obtained by conducting fully connected layer on each element of cell's output.

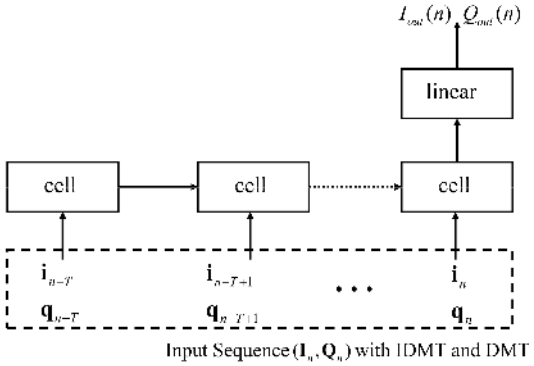



**FIGURE 5.** Sequence-to-one regression mode.

## V. ANALYSIS AND COMPARISON OF COMPUTATIONAL COMPLEXITY

In this part, we compare the complexity of different models. Firstly, the complexity and modeling performance of the two new proposed models are analyzed and compared. Secondly, we compare the complexity and modeling performance of the second model with GRU model.

### A. COMPLEXITY COMPARISON OF IGRNN AND IGIRNN

Suppose the number of hidden layers and its hidden neurons is 1 and G, respectively. The memory depth of the input vector is $M$. Since the input sequence length T does not bring change to the total parameters, the total coefficients to be updated in the IGRNN is:

$$N_1 = 3 \times 2 \times (M+1) \times G + G^2 + 5 \times G \tag{29}$$

The total coefficients to be updated in the IGIRNN is:

$$N_2 = 2 \times 2 \times (M+1) \times G + G^2 + 5 \times G \tag{30}$$

And it yields:

$$N_1 - N_2 = 2 \times (M+1) \times G \tag{31}$$

The visualization of the relationship between the $N_1 - N_2$ and its corresponding memory depth M and the number of hidden neurons are illustrated in Fig. 6. Taking $G = 100$ and $M = 15$ as an example, the total parameters of the IGRNN and IGIRNN models are 20100 and 16900, respectively. The corresponding modeling performance is listed in TABLE 1.

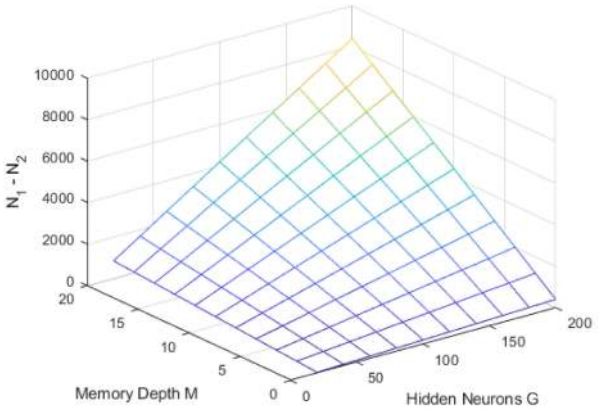



**FIGURE 6.** The relationship of (31).

**TABLE 1.** Performance comparison of modeling.

| Model Type | Training NMSE(dB) | Validation NMSE(dB) | Number of coefficients |
|---|---|---|---|
| GMP | -35.94 | -35.87 | 1052 |
| VDTDNN | -35.61 | -35.48 | 12202 |
| RNN | -35.27 | -35.03 | 13600 |
| GRU | -38.91 | -38.27 | 40100 |
| BiLSTM | -41.87 | -39.17 | 107600 |
| IGRNN | -38.58 | -37.79 | 20100 |
| IGIRNN | -38.61 | -37.93 | 16900 |

### B. COMPLEXITY COMPARISON OF IGIRNN AND GRU

At present, there are many variant structures based on RNN model, such as LSTM and GRU model. GRU has been proven to have good model expression ability in many fields, and it is easier to train with less parameters than LSTM model.

So here GRU is adopted as a comparison model to show that the IGIRNN model has the comparable performance with lower computational complexity. The total coefficients to be updated in the GRU is:

$$N_3 = 3 \times 2 \times (M+1) \times G + 3G^2 + 5 \times G \tag{32}$$

And it yields:

$$N_3 - N_2 = 2 \times (M+1) \times G + 2 \times G^2 \tag{33}$$

The visualization of the relationship between the $N_3 - N_2$ and its corresponding memory depth $M$ and the number of hidden neurons $G$ are illustrated in Fig. 7. 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 wideband signals. Therefore, with the increase of G, the parameter of IGIRNN will obviously decrease compared with the GRU model. Taking $G = 100$ and $M = 15$ as an example, the total parameters of the GRU and the IGIRNN models are 40100 and 16900, respectively. Also the corresponding modeling performance is listed in TABLE 1.

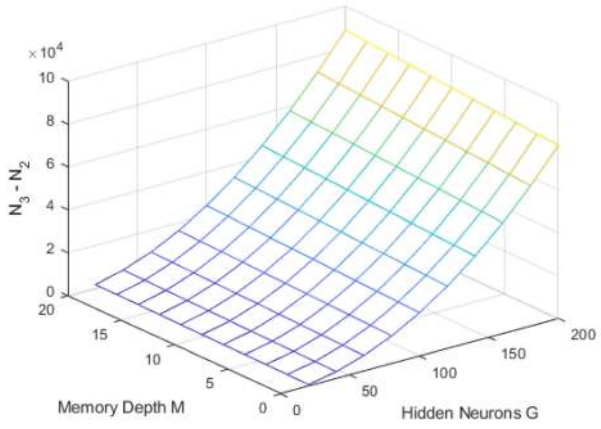



**FIGURE 7.** The relationship of (33).

## VI. EXPERIMENTAL VALIDATION

In order to compare the performance of the proposed models and the existing models, various experimental tests were conducted. The photograph of the experimental platform is shown in Fig. 8. The platform is composed of a personal computer (PC) with MATLB and pytorch softwares, a Vector Signal Generator (SMW200A) from R&S, a linear driver PA, a wideband Doherty PA operated at 2.4GHz, a 40 dB RF attenuator and a Spectrum Analyzer (FSW43) also from R&S. The flow diagram of the modeling procedure with Iterative Learning Control (ILC) is depicted in Fig. 9. Firstly the upconverted RF signals were generated by the Vector Signal Generator under the control of PC. Secondly, after the linear driver PA, the RF signals were amplified with the Doherty PA operated at 2.4GHz. Thirdly, the attenuated PA outputs were captured by the Spectrum Analyzer and sent into the PC. Finally, the parameters of the model are extracted and trained in PC(Intel Core i5-9400 CPU, 16G RAM).

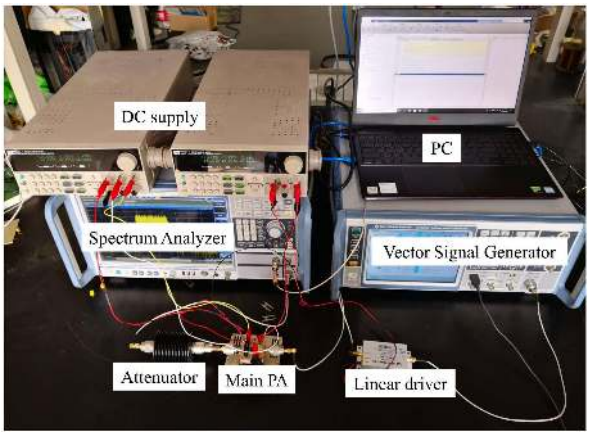



**FIGURE 8.** The photograph of the test bench.

The idea of ILC is to change the input of unknown system through training, so as to minimize the error between the output of actual system and the output of ideal system [25]. As shown in Fig. 9, ILC is applied to the parameter identification of DPD. Firstly, through ILC algorithm, the ideal PA input $\mathbf{z}_{ideal}$ can be obtained by iteration and this makes the error of PA output $\mathbf{y}$ and the digital predistorter input $\mathbf{x}$ minimum, that is, $\mathbf{z}_{ideal}$ is the optimal predistortion signal $\mathbf{z}_{opt}$. Then by choosing the appropriate model, the digital predistorter can be identified according to the input of the digital predistorter and the best predistortion signal.

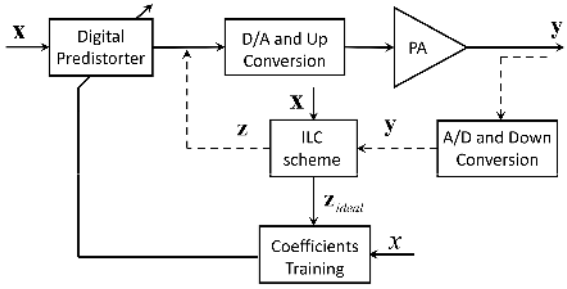



**FIGURE 9.** The flow diagram of the modeling procedure with ILC.

### A. THE PERFORMANCE COMPARISON OF MODELING

To evaluate the modeling ability of the proposed models, the conventional GMP model, the VDTDNN model, the RNN model, the GRU model and BiLSTM model are used for comparison. The test signal is a 5-carrier 100MHz OFDM signal and its PAPR is 6.6dB. Around 60,000 samples were recorded with the sampling rate at 368.64 MSPS. The first 50,000 samples were used to extract model coefficients and the remaining samples were used for performance verification.

The test signal and the PA output signal are then used to train the models. The configurations of different 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. The proposed models, RNN, GRU and BiLSTM here have 1 layer with 100 hidden neurons; The memory depths $M$ of the input vectors of the proposed models, RNN, GRU and BiLSTM are all 15, and the input sequence length of the proposed models and GRU model are 8. VDTDNN here has 2 layer with 100 hidden neurons; The memory depths $M$ of the input vectors of VDTDNN is15. It should be noted that the configurations of GMP here has already reached present the possible best performance, although more coefficients can be selected. The modeling performance, denoted by normalized mean square error (NMSE) for the digital predistorter of different model are listed in TABLE 1.

It is clear compared with GMP model, the VDTDNN and the RNN, the two proposed novel models can improve NMSE at least 1.92 dB for the validation performance. Therefore, in wideband scenarios, the two novel models are promising behavioral models for modeling digital predistorters and PAs. In addition, GRU and BiLSTM have better performance than proposed novel models, but those models complexity are higher.

For the sake of further comparison the modeling performance of different neural models, the number of hidden neurons are changed to evaluate the performance. The number of hidden neurons changes from 20 to 120 at an interval of 20. It can be found from Fig. 10 that the performance of traditional RNN structure modeling is poor. And with the increase of the number of neurons, the performance of the two proposed models are close to GRU and BiLSTM model. When the number of neurons is small, the performance of IGRNN model is better than that of IGIRNN model, but with the increase of the number of neurons, the performance of IGIRNN is close to or even better than IGRNN. The reason for this phenomenon may be that with the increase of neurons, the ability of network structure expression is enhanced, which may cause redundancy of input information.

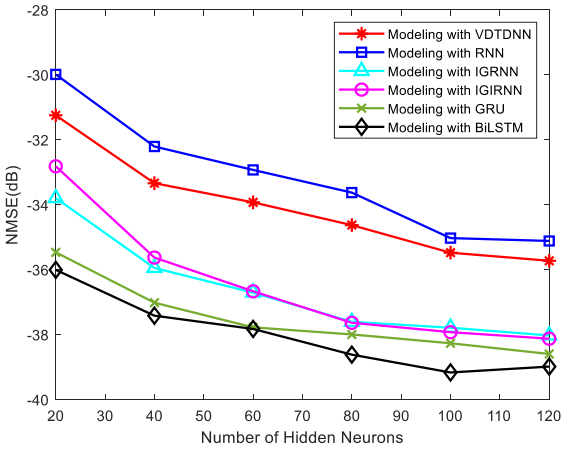



**FIGURE 10.** Performance comparison modeling with varied number of hidden neuron.

### *B. THE LINEARIZATION PERFORMANCE COMPARISON OF GMP, VDTDNN, RNN, IGRNN AND IGIRNN*

In part A of this section, the modeling performance of different models was compared. Here the linearization performance of different models is further compared. To make the comparison clearer, this part B only compares the linearization performance of GMP, VDTDNN, RNN and the two proposed models. In the next part, the linearization performance of GRU and the two proposed models will be compared. All the settings in this part are the same as part A.

The AM-AM and AM-PM characteristics without DPD and with IGRNN model DPD are shown in Fig. 11. The AM-AM and AM-PM characteristics without DPD reflects the working state and distortion of the RF-PA. The AM-AM and AM-PM characteristics with IGRNN model DPD directly reflects the linearization performance of IGRNN. The detailed NMSE and adjacent channel power ratio (ACPR) are listed in TABLE 2. The power spectral density (PSD) comparision is shown in Fig. 12.

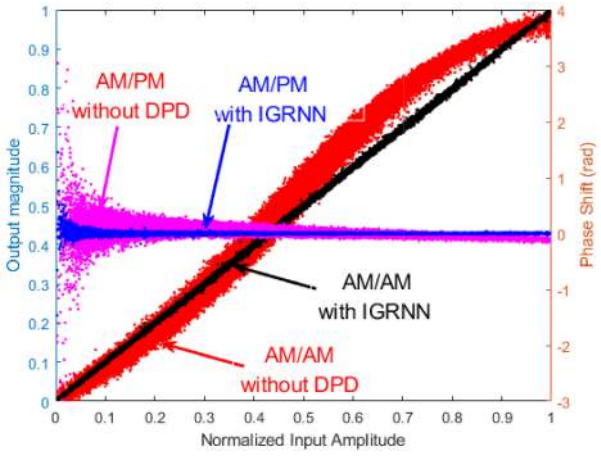



**FIGURE 11.** AM-AM and AM-PM plots with and without DPD for a 100-MHz OFDM signal.

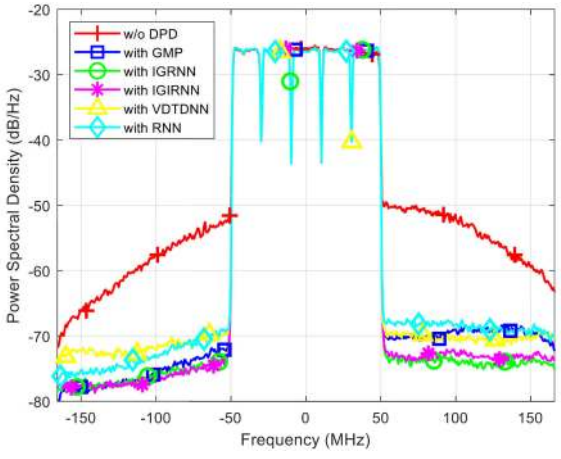



**FIGURE 12.** The power spectral density comparison of GMP, VDTDNN, RNN, IGRNN and IGIRNN for 100-MHz OFDM signal.

**TABLE 2. Performance comparison of GMP, VDTDNN, RNN, IGRNN and IGIRNN for 100-MHz OFDM signal.**

| Type of Model | NMSE (dB) | ACPR(dB) (lower/upper) | Training Time(s) |
|---|---|---|---|
| Without DPD | -19.70 | -26.38 /-23.46 | N/A |
| GMP | -38.47 | -45.95/-42.84 | 142 |
| VDTDNN | -37.59 | -43.30/-42.52 | 1272 |
| RNN | -37.21 | -43.37/-41.56 | 850 |
| IGRNN | -40.62 | -47.34/-46.35 | 1595 |
| IGIRNN | -41.09 | -47.68/-45.90 | 1368 |

According to TABLE 2 and Fig. 12, the NMSE of the IGRNN and IGIRNN can be improved at least 2.15dB compared with GMP, VDTDNN and RNN. The ACPR of the IGRNN and IGIRNN at the lower sideband can be improved at least 1.39dB, while the ACPR at the upper sideband can be improved at least 3.06dB. The performance of IGIRNN is equivalent to that of IGRNN, and at some frequencies the performance of IGIRNN can be even slightly better than that of IGRNN.

### C. THE LINEARIZATION PERFORMANCE COMPARISON OF GRU, IGRNN AND IGIRNN

In this part, the proposed models are compared with the RNN variant based gate structure model, of which GRU model is a typical one. In addition, it can be seen from part A that BiLSTM has similar modeling performance to GRU, but the complexity is too high, which is not practical in DPD experiment. Therefore in this part, the linearization performance of GRU and the two proposed models are compared. In this experiment, a 120MHz OFDM signal was employed as the test signal. Other experimental settings in this part are the same as those in part A.

It can be seen from TABLE 3 and Fig. 13, all three models have excellent predistortion performance. The linearization performance of the proposed two models is comparable to that of GRU model, and the number of parameters is significantly reduced. When the number of neurons is 100 and the memory depth is 15, the number of model parameters of IGRNN and IGIRNN relative to GRU reduces 49.9 percent and 57.8 percent, respectively. Besides, the characteistic curves of the AM-AM and AM-PM without DPD and with IGIRNN are shown in Fig. 14.

**TABLE 3. Performance comparison of GRU, IGRNN and IGIRNN for 120-MHz OFDM signal.**

| Type of Model | NMSE (dB) | ACPR(dB) (lower/upper) | Training Time(s) |
|---|---|---|---|
| Without DPD | -20.21 | -27.48/-25.56 | N/A |
| IGRNN | -39.34 | -44.74/-44.43 | 1621 |
| IGIRNN | -39.76 | -44.86/-44.56 | 1383 |
| GRU | -40.05 | -45.02/-44.82 | 2350 |

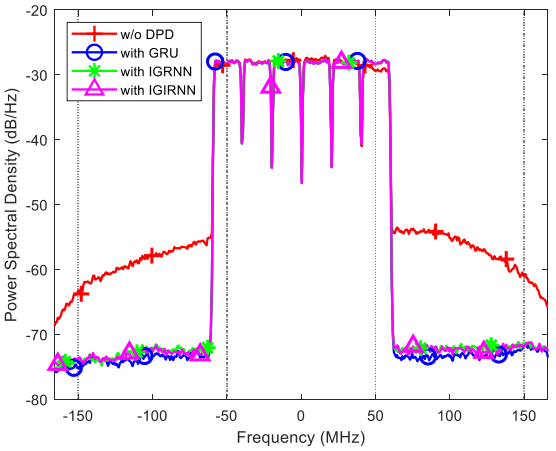



**FIGURE 13. The power spectral density comparison of GRU, IGRNN and IGIRNN for 120-MHz OFDM signal.**

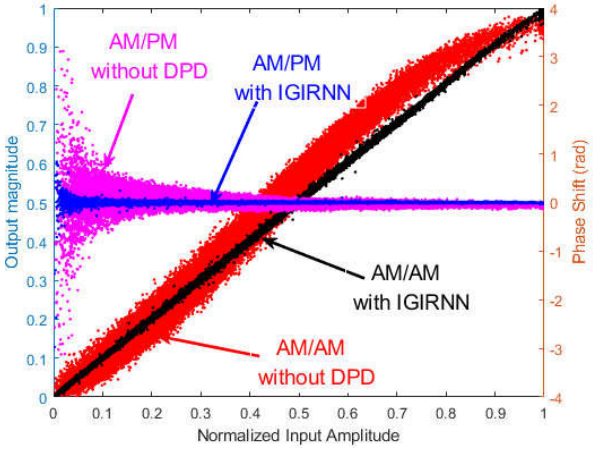



**FIGURE 14. AM-AM and AM-PM plots with and without DPD for a 120-MHz OFDM signal.**

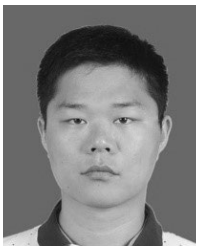
## VII. CONCLUSION

In this study, two novel neural network models based on the RNN model are proposed. Different from the traditional RNN model, the two models proposed solve the problem of vanishing gradients in RNN model training. Therefore, can better describe the short-term and long-term memory effect of RF-PA. Additionally they are more in line with the physical distortion characteristics of power amplifiers. Compared with other RNN-based variant models, such as GRU model, the two proposed models are simpler and can effectively reduce model parameters. The theoretical analysis and various experimental results imply that the proposed IGRNN and IGIRNN have excellent performance compared with traditional DPD models such as GMP model, and significantly lower computational complexity compared with RNN variant models such as GRU model.

## REFERENCES

[1] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 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] S. Kang, E.-T. Sung, and S. Hong, "Dynamic feedback linearizer of RF CMOS power amplifier," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 915–917, Oct. 2018.
[4] A. Katz, J. Wood, and D. Chokola, "The evolution of PA linearization: From classic feedforward and feedback through analog and digital predistortion," *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 32–40, Feb. 2016.
[5] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.
[6] 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.
[7] 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.
[8] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.
[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] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*. Norwood, MA, USA: Artech House, 2006.
[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[12] 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.
[13] 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.
[14] D. Luongvinh and Y. Kwon, "Behavioral modeling of power amplifiers using fully recurrent neural networks," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2005, p. 4.
[15] 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.
[16] J. Sun, W. Shi, Z. Yang, J. Yang, and G. Gui, "Behavioral modeling and linearization of wideband RF power amplifiers using BiLSTM networks for 5G wireless systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10348–10356, Nov. 2019.
[17] 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.
[18] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[19] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. Int. Conf. Mach. learn.*, 2013, pp. 1310–1318.
[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[21] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," in *Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN)*, vol. 2, Sep. 1999, pp. 850–855.
[22] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078
[23] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.
[24] P. Roblin, D. E. Root, J. Verspecht, Y. Ko, and J. P. Teyssier, "New trends for the nonlinear measurement and modeling of high-power RF transistors and amplifiers with memory effects," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1964–1978, Jun. 2012.
[25] 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.



**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 the nonlinear modeling of transmitters.

**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 was a Visiting Ph.D. Student with the IoE2 Laboratory, School of Electrical and Electronic Engineering, University College Dublin, Dublin, Ireland. His research interests include digital predistortion linearization, nonlinear system identification algorithms, and machine learning.

**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 the digital predistortion of power amplifiers, nonlinear system identification, and machine learning.

**WEN QIAO** received the B.E. degree from the School of Electronics and Information Engineering, Anhui University, Hefei, China, in 2016. She is currently pursuing the Ph.D. degree in electromagnetic filed and microwave technology with the University of Science and Technology of China (USTC). Her research interests focus on digital predistortion and the 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).