IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 February 2023, accepted 14 March 2023, date of publication 17 March 2023, date of current version 22 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3258691

RESEARCH ARTICLE

Accurate and Effective Nonlinear Behavioral Modeling of a 10-W GaN HEMT Based on LSTM Neural Networks

MINGQIANG GENG[1], (Student Member, IEEE), GIOVANNI CRUPI[2], (Senior Member, IEEE), AND JIALIN CAI[1], (Senior Member, IEEE)

[1]Key Laboratory of RF Circuit and System, Ministry of Education, College of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, China
[2]BIOMORF Department, University of Messina, 98125 Messina, Italy

Corresponding author: Jialin Cai (caijialin@hdu.edu.cn)

This work was supported in part by the Provincial Universities of Zhejiang under Grant GK229909299001-011, in part by the Qianjiang Talent Project Type-D of Zhejiang under Grant QJD2002020, and in part by the National Natural Science Foundation of China (NSFC) under Grant 61971170.

ABSTRACT This paper presents a novel nonlinear behavioral modeling methodology based on long-short-term memory (LSTM) networks for gallium nitride (GaN) high-electron-mobility transistors (HEMTs). There are both theoretical foundations and practical implementations of the modeling procedure provided in this paper. To determine the most appropriate optimizer algorithm for the model presented in this work, four different optimization algorithms are examined. The results of both simulation and experimental validation are provided based on a 10-W GaN HEMT device. According to the developed investigation, the model is capable of extrapolating and interpolating over multiple input power levels and frequencies, including linear, weakly nonlinear, and strongly nonlinear areas. The analysis of the simulated and measured results shows that the developed model has superior performance also when considering the DC drain current (*Ids*.). Compared with the existing support vector regression (SVR) based model and the Bayesian based model, the proposed approach shows a significantly improved extrapolation capability.

INDEX TERMS Behavioral modeling, black box model, gallium nitride (GaN), high-electron-mobility transistor (HEMT), long-short term memory (LSTM), microwave frequency, power transistor.

I. INTRODUCTION

GaN HEMTs have recently emerged as one of the most popular devices for front-end applications in radio frequency (RF). In contrast to conventional semiconductor materials, e.g., gallium arsenide (GaAs), GaN devices are capable of operating at higher critical electric fields, exhibit broader band gaps, and achieve higher electron mobility [1], [2], resulting in a substantial increase in the magnitude of power density [3]. Thus, it has become the dominant power transistor used in the design of power amplifiers (PAs) [4], [5], [6], [7], [8], [9], allowing outstanding performance to be achieved.

In spite of these significant advantages, further research is necessary to understand the inner nonlinear electrical characteristics of GaN HEMT devices, particularly their strong nonlinear effects and severe dispersive effects [10], [11], [12], [13]. It has been reported that these phenomena can negatively affect the efficiency and output power of the device. This can lead to a degradation of performance for the whole circuit, or even for the entire system [14]. In this regard, the development of an accurate model for GaN HEMT devices is urgently needed, but also a very challenging task.

As a result, various modeling approaches have been developed in recent years [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], including physical base models [15], equivalent circuit models [16], [17], [18], behavioral models [19], [20], [21], [22], [23] and combinatorial models [24], [25], [26]. The present work focuses on modeling GaN HEMT technology using a behavior-based approach. The advantage of this choice is that, firstly, the behavioral

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

modeling method, as an alternative to the other approaches, provides a more accurate prediction of the working regions used for extraction. Furthermore, as a complete "black box" model, it takes into account only the input and output information of the device under test (DUT) [19], [20], [21], [22], [23].

Numerous studies have been conducted on the behavioral modeling of RF power devices. In addition, several books have been published that summarize the main contributions to this challenging and stimulating field of research [27], [28]. Behavioral modeling techniques for power devices have been briefly reviewed in [29], which provides an overview of this active area of research. Behavioral models based on traditional approximations [19], [21], [30], [31], such as polynomials, produce reasonable predictions, however these models are not able to interpolate and extrapolate well for strong nonlinear cases. In comparison to traditional behavioral modeling methods, artificial neural network (ANN) modeling techniques offer an effective alternative [32], [33], [34], [35]. Machine learning (ML) techniques, which are a core component of artificial intelligence, have been shown to provide superior results in comparison to neural networks [36], [37], [38]. In [37], a device modeling algorithm based on Bayesian inference was developed that can predict the fundamental scattered waves of power transistors at a variety of input power levels. The model is capable of excellent interpolation. Small-signal and large-signal behavioral models based on support vector regression (SVR) were implemented in [26] and [38], and both showed accurate prediction performance.

While all of these models, whether they are traditional approximation models or machine learning models, are capable of providing accurate local and interpolated predictions, they are limited in their ability to extrapolate beyond the measurement range. It is critical to note that extrapolation is one of the most significant generalization capabilities of a behavioral model, particularly when dealing with nonlinear operating regions. Models with strong extrapolation capabilities reduce the complexity of measurement significantly, and this is particularly helpful for nonlinear behavioral modeling of GaN HEMT transistors, as it is more likely for the device to burn when tested under large mismatch conditions during load-pull measurements. In this work, an extrapolation-enhanced nonlinear behavioral model for GaN HEMT technology is proposed and validated.

In [39], a modified recurrent neural network (RNN) technique based on long-short term memory (LSTM) was proposed for the analysis of GaN HEMT small-signal behavior. According to [39], the model can accurately predict the small-signal behavior of the DUT. Further, the proposed LSTM technique is capable of representing both linear and nonlinear relationships between complex inputs and outputs, not only for single-input and single-output (SISO) problems but also for multiple-input and multiple-output (MIMO) problems. As a result of the properties of the LSTM architecture, it can overcome the issue of long-term dependencies, such as vanishing or exploding gradients [40].

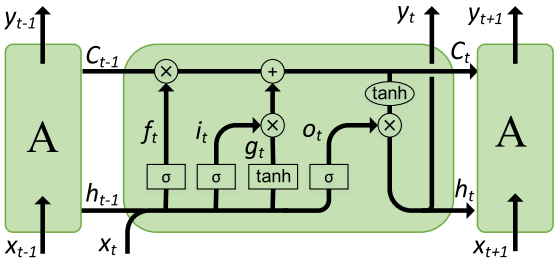


FIGURE 1. General structure of the LSTM networks and internal architecture of a single memory cell. The symbols of "⊕" and "⊕" represent the operation of addition and element-wise multiplication, while σ and tanh denote the activation function.

As a follow-up to [39], this paper employs the LSTM network to build a nonlinear behavioral model for GaN devices, including both RF and DC parts. To determine the optimal core function of the model for a given DUT, extensive tests were conducted. This model is shown to be capable of extrapolating and interpolating for three different input power regions and operating frequencies.

The paper is structured as follows. Section II describes the theory behind LSTM neural networks, along with the different optimizer algorithms, different topologies, and extraction details. Furthermore, the application of the LSTM method to model the behavior of a 10-W GaN HEMT device is described in detail. In Section III, both simulation and experimental verification are presented and discussed. In Section IV, the conclusions are summarized.

II. BASIC THEORY OF MODELING METHODOLOGY

A. THEORY OF LSTM NEURAL NETWORKS

LSTM is a widely used machine learning technique that originated from RNN and was first developed by Hochreiter and Schmidhuber [40]. According to [40], [41], and [42], RNN is not capable of capturing future information and is likely to encounter exploding or disappearing gradients when modeling large series of data, usually resulting in poor quality models. Additionally, the RNN gradient is dominated by the near-range gradient, making it difficult for the model to learn long-range relationships. By incorporating multiple gate controllers into memory cells, LSTM networks demonstrate a valuable and promising capability of solving these issues. Multiplying and adding the elements enables the network to store large amounts of previous data, allowing for long-term modeling, leading to a strong capability for extrapolation and interpolation.

The cell state of the LSTM networks is illustrated in Fig. 1, where the green diagram represents a basic LSTM memory cell. A complete LSTM neural network consists of several repeating unit modules, which can be represented as a chain structure. This structure shows the LSTM in the form of a time expansion. As can be seen in Fig. 1, there is one input x_t and the corresponding output y_t. The sequence-based model presents the specific connectivity of each LSTM unit,

which is able to establish the temporal association between the previous cell state C_{t-1} and the current cell state C_t, while transferring the output h_{t-1}. Similarly, the next cell information can be transferred in the same way.

Each cell consists of three gates, which are input, forget, and output gates. The LSTM updates the memory states, i.e., long-term state C_t and short-term state h_t, through the gate controllers [43]. First, the current input x_t and the previous short-term state h_{t-1} are used to produce the four gate parameters, f_t, i_t, g_t, and o_t. The cell output at the previous moment and the input at the current moment are used to determine which vectors of the internal state should be updated or maintained. The signal of the forget gate controls how much of the previous state information should be forgotten, while the input gate reacts to the effect of input x_t in the current cell state, and finally, the output gate controls y_t or h_t as the output of the LSTM, where the output y_t is equal to h_t.

The three gates in a memory cell are given by the following set of formulations. The first forget gate contains a sigmoid control layer and it can be expressed as:

$$f_t = \sigma(w_{xf}^T \cdot x_t + w_{hf}^T \cdot h_{t-1} + b_f) \tag{1}$$

where f_t is the output of the forget gate, w_{hf} and w_{xf} are denoted as the weight matrices for the previous short-term state h_{t-1} and input x_t, respectively, and b_f is the bias terms.

The second gate is the input gate. They can be expressed as:

$$i_t = \sigma(w_{xi}^T \cdot x_t + w_{hi}^T \cdot h_{t-1} + b_i) \tag{2}$$

$$g_t = tanh(w_{xg}^T \cdot x_t + w_{hg}^T \cdot h_{t-1} + b_g) \tag{3}$$

where i_t is the output of the input gate, g_t is candidate status, w_{hi} and w_{hg} are denoted as the weight matrices for the previous short-term state h_{t-1}, w_{xi}, w_{xg} are weight matrices for input x_t, and, similarly, the bias terms are given by b_i and b_g.

Thirdly, the output gate function is given as:

$$o_t = \sigma(w_{xo}^T \cdot x_t + w_{ho}^T \cdot h_{t-1} + b_o) \tag{4}$$

where o_t is the output of the output gate. Similarly, each of the other coefficients in the equation indicates the same meaning as in the gate above (see eq. (3)).

Finally, after completing the calculation within the intra-cell, the information of the memory states is updated. This can be described as follows:

$$C_t = f_t \otimes C_{t-1} + i_t \otimes g_t \tag{5}$$

$$y_t = h_t = o_t \otimes tanh(C_t) \tag{6}$$

The sigmoid layer of the input gate determines which information has to be updated, the tanh layer is used to create a new candidate value that might be added to the states of the cell. In this way, the LSTM networks can store large amounts of previous information, thus making it simple to learn long-term relationships.

B. BEHAVIORAL MODELING FOR RF TRANSISTORS

For the nonlinear behavioral modeling of a power transistor, the behavior of the DUT can be described by the incident and scattered waves A_{qn} and B_{pm}. Typically, the scattered wave response is obtained by apply excitation, i.e., A_{qn}, to the device. Mathematically, these waves can be expressed as follows:

$$A_{qn} = \frac{V_{qn} + Z_0 I_{qn}}{2\sqrt{Z}_0} \tag{7a}$$

$$B_{pm} = \frac{V_{pm} - Z_0 I_{pm}}{2\sqrt{Z}_0} \tag{7b}$$

where Z_0 is the characteristic impedance taken here to be real-valued, and, for this article, taken as 50 Ω, q and p represent the respective incident and scattered wave ports, and n and m range from zero (DC) to the highest harmonic index at the incident and scattered ports, respectively.

For a given DUT, the describing function f_{pm}, as given in eq. (8), can associate all of the relevant incident waves A_{qn} with the scattered waves B_{pm}:

$$B_{pm} = f_{pm}\left(A_{11}, A_{12}, \ldots, A_{1N}, A_{21}, A_{22}, \ldots, A_{2M}\right) \tag{8}$$

where A_{11},A_{12},…,A_{21},A_{22},…represent all of the incident waves. The independent variable of the describing function f_{pm} is the incident wave phasors A_{qn} at both the input and output ports at the fundamental frequency f_0, and the corresponding output is the scattered wave phasors B_{pm}. The aim of modeling the DUT becomes replacing the describing function, f_{pm}, by a proposed large-signal model equation.

Refer to eq. (8), the inputs of the descriptive function are complex-valued, whereas the proposed LSTM modeling technique is suited for real-valued cases. Thus, a modification of the standard LSTM method is required. In [44], a real-valued time-domain ANN modeling technique is presented and, in [35], it was well adapted to the frequency-domain modeling methods. Thus, a similar decompose technique is employed for the LSTM modeling method. The description formulation of the scattered wave is split and expressed as a function of both the real and imaginary parts of the incident waves and, then, the specific equation is rewritten as follows:

$$B_{pm}^R = f_{pm}^R\left(\overbrace{A_{qn}^R, A_{qn}^I}\right) \tag{9a}$$

$$B_{pm}^I = f_{pm}^I\left(\overbrace{A_{qn}^R, A_{qn}^I}\right) \tag{9b}$$

where B_{pm}^R and A_{qn}^R denote the real parts of the incident and scattered wave phasor, and B_{pm}^I and A_{qn}^I are the imaginary parts of the incident and scattered wave phasor, respectively. The superscript in the marker represents the set of all real or imaginary parts of the wave, for each harmonic index.

C. LSTM TECHNIQUE BASED BEHAVIORAL MODEL

Based on the characteristics of the LSTM modeling technique, the complete process of behavioral modeling of GaN

device is implemented in *TensorFlow* using the LSTM neural networks. The detailed steps of the modeling procedure are given below:

1) Initialize the design variables, set the iteration number of epochs and batch sizes, and import the training data.
2) Build the LSTM neural networks. One layer of the LSTM network, with several layers of Dense layers, was applied in this work.
3) Select the mean square error (MSE) as the loss function.
4) Specify global optimizers with a learning rate. Several mainstream optimizers have been selected and examined in this work and a detailed discussion will be given later.
5) Feed the training data into the network architecture. Depending on the characteristics of the data, a multi-dimensional to one-dimensional input-output model is used.
6) When the iteration number is reached, the behavioral model can be obtained and used to predict new output.

In order to choose the optimal optimizer in the proposed model, four different popular optimizers have been selected and compared, which are the stochastic gradient descent (SGD) method, the root mean square prop (RMSprop) method, the adaptive moment estimation (Adam) method, and the Nesterov Adam (Nadam) method.

SGD was shown to be an effective optimization method for minimizing the cost function $J(\theta)$ by updating the parameters in the opposite direction of the cost function gradient with model parameter θ [45] and it has been successfully used in many machine learning modeling cases [41], [45], [46]. RMSprop was proposed by Tieleman and Hinton [47], which is an adaptive learning rate method. *Adam* is an algorithm for first-order gradient-based optimization of stochastic functions, based on adaptive estimates of lower-order moments [48]. It is a versatile algorithm that can scale to large-scale high-dimensional machine learning problems. *Nadam* was proposed by Dozat [49]. It combines *Adam* with Nesterov momentum, which has a strong constraint on the learning rate and has a more direct impact on the gradient update [50].

After introducing four different optimizers, the LSTM technique is employed for device modeling. The detailed topology, for nonlinear behavioral modeling, is given in Fig. 2. The incident and scattered waves used for training are divided into real and imaginary parts and then imported into the model. Both machines get the same input, but the outputs of the first and second machines are the real and imaginary parts of the scattered waves, B_{pm}, respectively. It is critical to note that the data used for training and testing are different.

III. MODEL VALIDATION

In this section, model verification is carried out by using both simulations and experimental tests. All the LSTM based models are extracted using Python programing language.

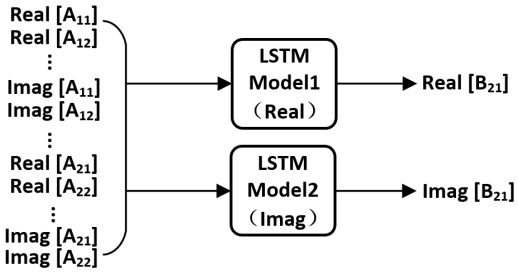


FIGURE 2. Block diagram of large-signal behavioral model based on LSTM networks, which consists of two LSTM machines. All wave phasor has been divided into real and imaginary parts, respectively.

TABLE 1. Performance comparison of LSTM model with different optimizers.

Model	Average Relative Error (%)	Extraction Time (s)	Learning Rate
SGD	2.95	7.43	0.01
RMSprop	0.28	12.32	0.001
Adam	0.17	14.76	0.001
Nadam	0.19	14.95	0.001

A. SIMULATIONS

A 10-W GaN packaged transistor (CGH40010F) from Wolfspeed was used in the simulation tests and an equivalent-circuit model from the same manufacturer was employed for data collection.

In the first step of the validation, the four different optimizers mentioned before of the LSTM are tested, the modeling performance is compared to fix the optimal optimizer for the proposed modeling method. The LSTM model used in the simulation is extracted with the device biased at 28 V for V_{DS} and -3 V for V_{GS}, with the operating frequency set at 1 GHz. Totally 80 load-pull sample points, uniformly distributed over the Smith chart, are used for training, when input power was set at $+25$ dBm. The second- and third-harmonic load impedances are fixed at 50Ω, which ensures that the second and third harmonic incident waves at port 2, A_{22} and A_{23}, are zero.

In the test, a one-layer LSTM network is used, with a batch size of 32 for the input data, and two layers of Dense for transforming the dimensionality of the data are employed, while, the output dimensionalities of the first- and second-Dense layer are fixed at 25 and 1, respectively. Four different optimizer algorithms: SGD, RMSprop, Adam, Nadam are employed, in turn, in the LSTM model. The prediction results of the four different LSTM models, compared to the reference circuit model, are provided in Fig. 3. As shown in this figure, the SGD optimizer has the worst prediction in this case, since the predicted points deviate significantly from the circuit data, while all the other three optimizers give an excellent prediction. A detailed comparison is shown in Tab. 1. All extraction and simulation times are for a PC with Intel Core i5-10500 Dual CPU 3.10 GHz and 24.00 GB of RAM, same

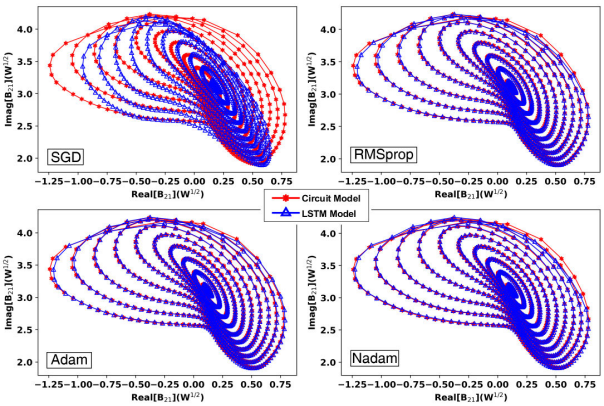


FIGURE 3. Performance test for LSTM behavioral model using four different optimizers, with +25 dBm available input power at 1 GHz.

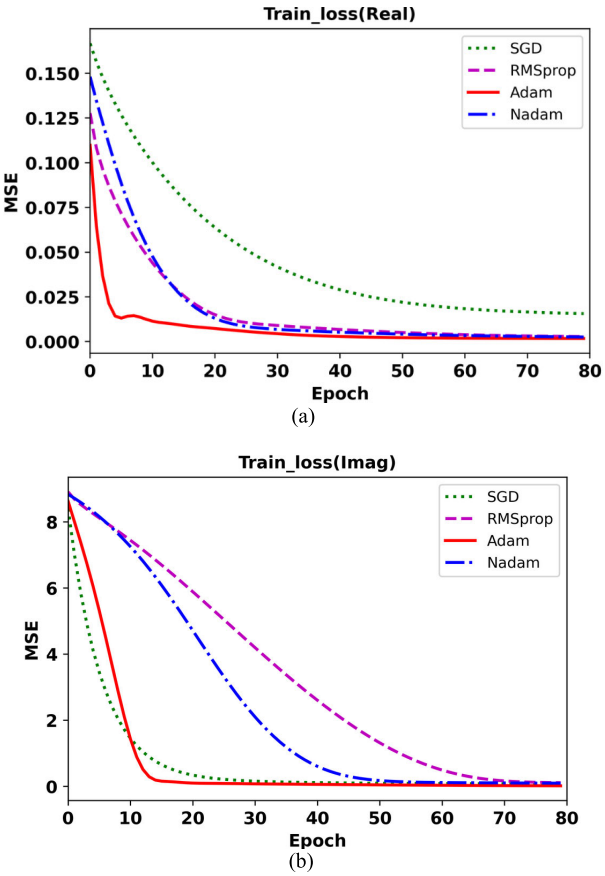


FIGURE 4. (a) Real part training loss and (b) imaginary part training loss of different optimizers on the LSTM model.

for the rest of the content in this work. The model with SGD optimizer has a lower prediction accuracy, despite its model extraction time is much shorter than the other models. The remaining three models demonstrate excellent accuracy with an average relative error much less than 1%. The expression of relative and average errors is show at the bottom of the next page. The models with Adam and Nadam optimizers have even better accuracy than the model with RMSprop optimizer, with the average errors here less than 0.2%. Compared with the Nadam, the Adam optimizer requires less time for extraction and provides better accuracy at the same time.

Loss functions of all the four training models are also illustrated in Fig. 4. Obviously, the performance of the SGD optimizer for the real and imaginary parts varies dramatically, and the *Nadam* converges slightly faster than RMSprop. *Adam*, on the other hand, consistently performs well and converges faster than the other three optimizers. Thus, in this work, the *Adam* algorithm is chosen and employed as the optimizer for the LSTM behavioral model of the DUT in the remaining tests of this paper.

In the second simulation test, the same DUT was stimulated by +20 dBm input at 2.5 GHz, with the same bias. The same 80 sample points were used for model extraction. The second- and third-harmonic load reflection coefficients are fixed at $\Gamma_{22} = 0.2/100$ and $\Gamma_{23} = 0.7/200$, respectively. To cover the whole Smith chart, validation is performed at all 720 load points. In the test, the performance of the proposed LSTM model is compared with the existing SVR-based model [51], the Bayesian inference-based model [37], and a multi-layer perception (MLP) based feed-forward neural network (FNN) model. The radio basis function (RBF) is employed as the kernel for the SVR model, and the grid search method is utilized to optimize the parameters. In the Bayesian model, the Gaussian function is used as the kernel function. While for the ANN-based model, a two-layer MLP with 30 neurons is used. The performance of the four models is shown in Fig. 5. It is observed that the fundamental scattered wave of the first three models matches the circuit model

TABLE 2. Performances comparison of different models.

Model	Extraction Time (s)	Average Relative Error (%)
SVR	67.22	0.124
Bayesian	11.95	0.098
LSTM	12.57	0.094
MLP	0.76	1.625

fairly well throughout the Smith chart. This demonstrates the excellent model interpolation capabilities of A_{21}. The MLP model gives accurate predictions for most of the points, however, there are large errors in the strong nonlinear area. A detailed comparison of all models is given in Tab. 2. The first three models provide excellent prediction accuracy, and their relative errors are much less than 1%. The SVR model requires much more model extraction time than the other three models, mainly due to that it takes time to find the optimal hyperparameters. The MLP model has the shortest extraction time, but the lowest model accuracy.

The fundamental output power contour prediction performance of the proposed model is given in Fig. 6. The LSTM model is able to determine the maximum output power region in the Smith chart precisely, which is of the utmost importance to assist in circuit design. In addition, power added

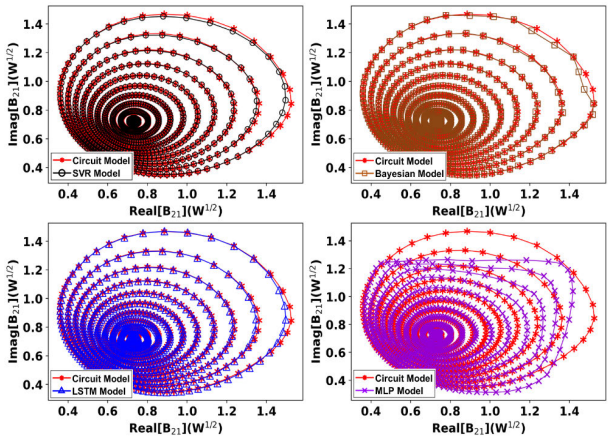


FIGURE 5. Simulation results from the circuit model, the SVR model, the Bayesian model, the MLP model, and the LSTM model, with +20 dBm available input power at 2.5 GHz. The second- and third-harmonic load reflection coefficients are fixed at Γ 22=0.2/100 and Γ 23=0.7/200, respectively.

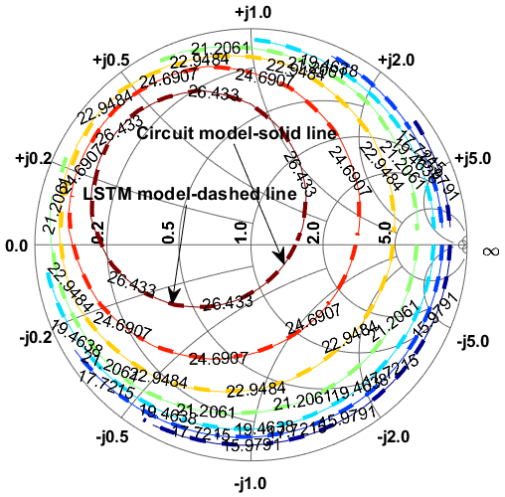


FIGURE 6. Fundamental output power contour with +20 dBm available input power at 2.5 GHz.

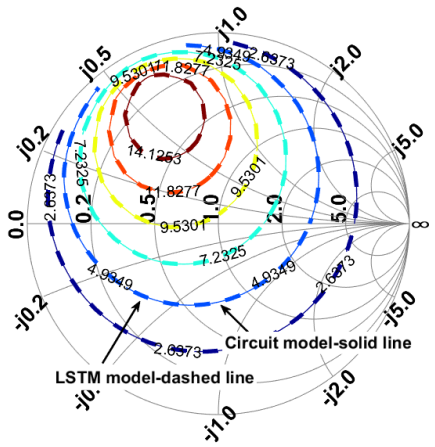


FIGURE 7. Power added efficiency contour with +20 dBm available input power at 2.5 GHz.

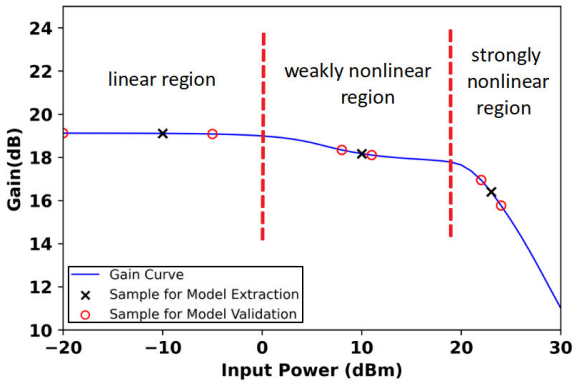


FIGURE 8. Gain curve versus input power (blue line), showing three different regions: linear, weakly nonlinear, and strongly nonlinear regions. Illustration of the samples used for model extraction (black crosses) and for the proposed model testing (red circles).

efficiency (PAE), an imperative factor for PA design, is also presented in Fig. 7. The proposed model can also accurately predict the PAE contour.

In the third simulation test, the extrapolation capability of the proposed LSTM model for fundamental scattered wave prediction is validated. The detailed test details are given below.

Fig. 8 gives a wide range of gain curves, with the input power varying from −20 dBm to +30 dBm and the operating frequency set at 1 GHz. The device bias and load samples used for model extraction are the same as in the previous test cases. Three different LSTM models were extracted at the following input power levels: -10 dBm, +10 dBm, and +23 dBm, which are located in three different regions of the device: the linear region, the weakly nonlinear region, and the strongly nonlinear region, as described in Fig. 8 with black crosses. Once model extraction is complete, the ability of the model to extrapolate between different input available power levels, for both A_{11} and A_{21}, can be verified. The test input power points are shown in Fig. 8 as well, in red circles.

In the linear region, where the model is extracted at −10 dBm, two input power points are tested, which

$$\text{relative error} = \frac{|measured\ value - modeled value|}{|measured\ value|} \times 100\%.$$

$$\text{average relative error} = \frac{\sum relative\ error at\ each\ load\ point}{number\ of load\ points}.$$

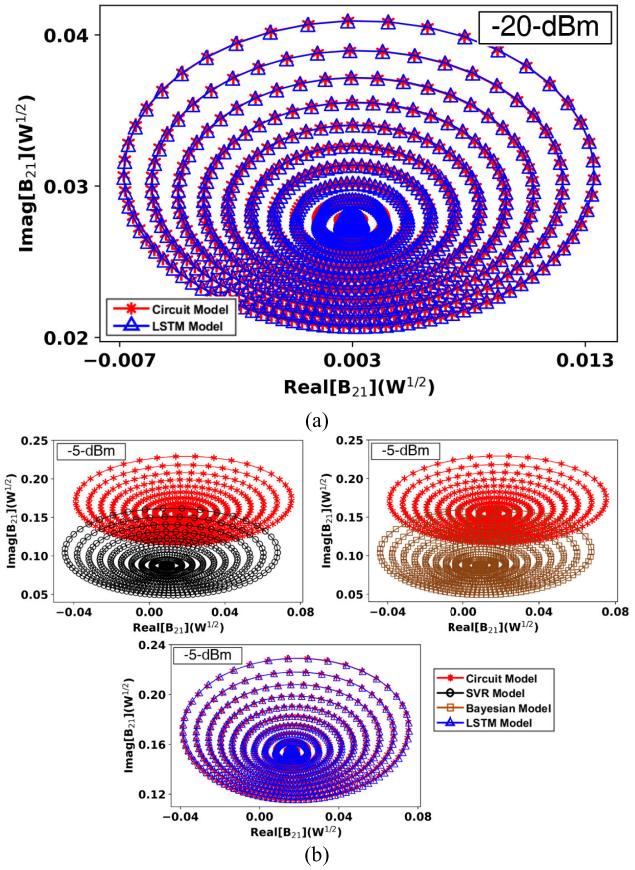


FIGURE 9. Extrapolation simulation results from the SVR model, the Bayesian model, and LSTM model with −20 dBm and −5 dBm input. All models are extracted with −10 dBm input.

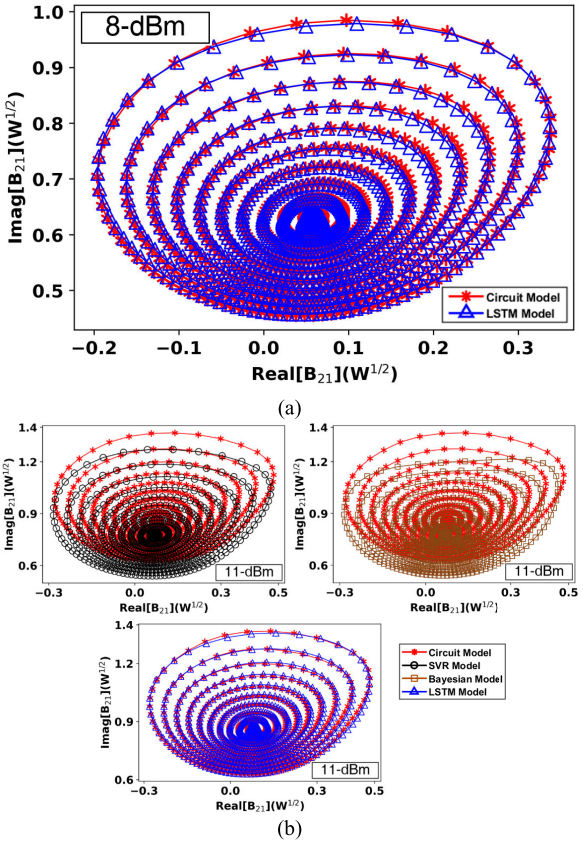


FIGURE 10. Extrapolation simulation results from the SVR model, the Bayesian model, and LSTM model with +8 dBm and +11 dBm input. All models are extracted with +10 dBm input.

are −20 dBm and −5 dBm. The results are shown in Fig. 9(a) and 9(b), respectively. The prediction performance of the SVR model and the Bayesian model for the -5 dBm case is also given in Fig. 9(b), by using black circles and brown squares. As can be seen, the proposed LSTM model can accurately match the fundamental scattered waves load-pull results at both two different input power cases, when the input power levels are far from the power used for extraction. Both the Bayesian and SVR models' predictions have a big shift, indicating that the proposed model has a much better extrapolation capability than the other two.

In the weakly nonlinear region, the model is extracted when the input power is +10 dBm. Two test input power levels are given, as shown in Fig. 8, with red circles. These test points are located 1 dB (+11 dBm), and 2 dB (+8 dBm) away from the model extraction points. The model extrapolation performance is illustrated in Fig. 10, as can be seen, the proposed model gives a prediction with high accuracy, for both different cases. The reference prediction given by the SVR model and the Bayesian model for the 11 dBm case shows a big shift, the same as in the previous linear case.

In the last test case for extrapolation capability, the test models were extracted at an input power of +23 dBm. This is located in the strongly nonlinear region, namely more than 3 dB compression, as shown in Fig. 8. The input power of the two test cases are ±1 dB away from the extraction point, which is +22 dBm and +24 dBm. The performance of all three models is presented in Fig. 11. As illustrated in the results, although the prediction of the LSTM model has a slight shift from the circuit results, it can still give reasonably accurate predictions for both input power levels. On the other hand, the SVR model shows a much larger deviation than the proposed one. In addition, the Bayesian model has the poorest prediction, for the +24 dBm case in Fig. 11.

From the three extrapolation tests above, the proposed model shows excellent extrapolation capability across all linear, weakly nonlinear, and strong nonlinear regions. The extrapolation performance of the LSTM model is also compared with the existing SVR model and Bayesian model. A big improvement was obtained for the proposed model in all three different test cases.

In the fourth simulation example, the extrapolation capability of the proposed LSTM model for DC drain current prediction at different input power levels is validated. The model structure is shown in Fig. 12. The same DUT is used, with the same bias, and the simulation data is identical to that in the third test case. However, the difference is that the

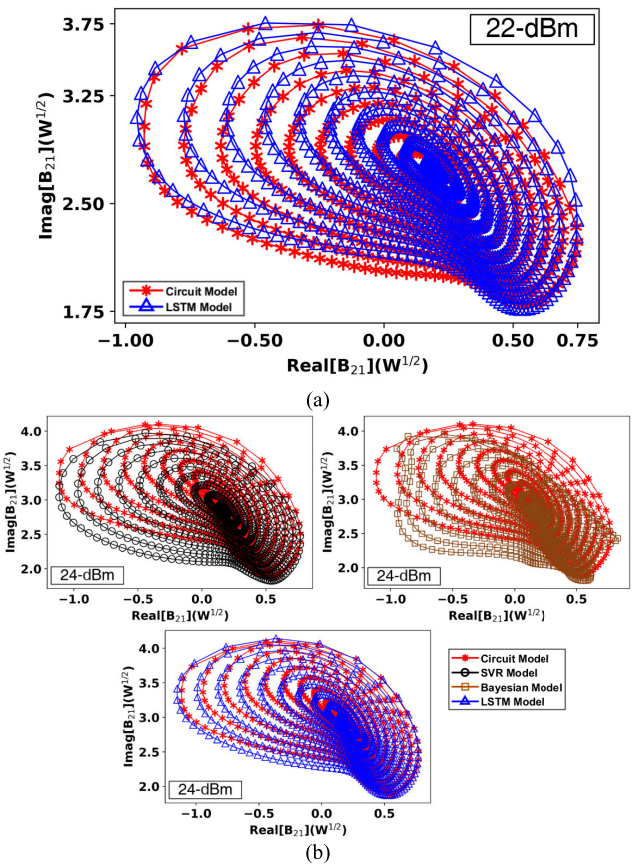


FIGURE 11. Extrapolation simulation results from the SVR model, the Bayesian model, and LSTM model with +22 dBm and +24 dBm input. All models are extracted with +23 dBm input.

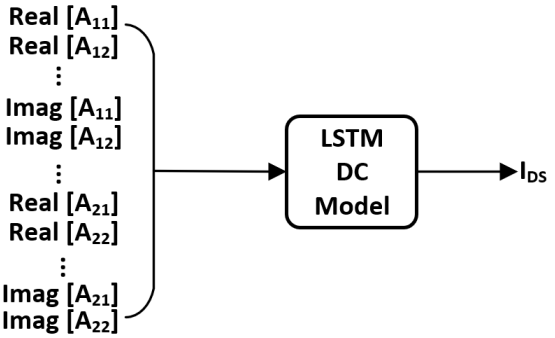


FIGURE 12. Block diagram of DC model based on LSTM networks. All wave phasor has been divided into real and imaginary parts, respectively.

model employs only a single LSTM machine to predict the output of the drain current *Ids*. The 720 points used for test validation cover the entire Smith chart. They sort these points sequentially by indexing them from the inner part to the outer part of the Smith chart.

Similar to the fundamental case, the DC model is extracted at three different input power levels, −10 dBm, +10 dBm, and +23 dBm, while the extrapolation ability is tested at levels of −5 dBm, +8 dBm, +11 dBm, and +24 dBm. The

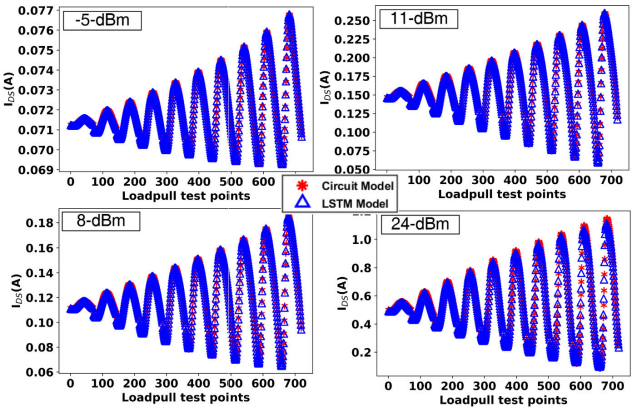


FIGURE 13. Comparison of the DC drain current results between the circuit model and the LSTM model at −5 dBm, +8 dBm, +11 dBm, and +24 dBm input with the operating frequency set at 1 GHz.

TABLE 3. Average relative error of LSTM DC current model for I_{DS}.

Model Extracted Input Power (dBm)	Model Tested Input Power (dBm)	Average Relative Error (%)
-10	-5	0.022
+10	+8	0.429
+10	+11	0.678
+23	+24	0.921

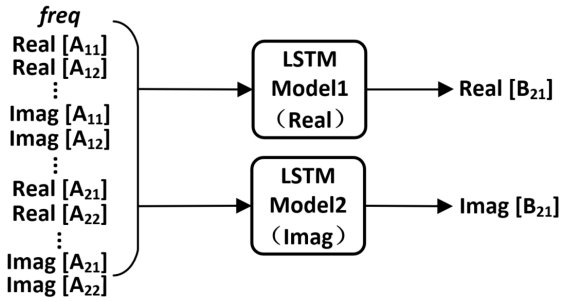


FIGURE 14. The topology of LSTM model with frequency information.

extrapolation prediction results for DC current are given in Fig. 13. As indicated in the figure, the proposed LSTM model can accurately predict every DC drain current point over a wide range of input powers. Further model details are shown in Tab. 3. The average relative error of the model is less than 1% over the entire test input power range, which validates the successful extrapolation performance of the proposed model in DC current prediction.

In the last simulation example, the interpolation ability of the proposed model at different input available power levels and operating frequencies was examined. For the power interpolation case, the model topology is the same as in Fig. 2, while for the frequency case, the operating frequency information needs to be added, thus, the updated topology is given in Fig. 14.

In the power interpolation test, the DUT was biased at 28 V for VDS and −3 V for VGS. The model was extracted when

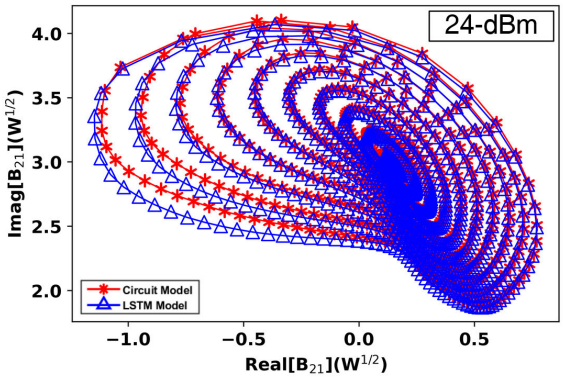


FIGURE 15. Input power interpolation test of the proposed LSTM model (blue triangles) at +24 dBm, and model extracted with +23 dBm and +25 dBm input at 1 GHz.

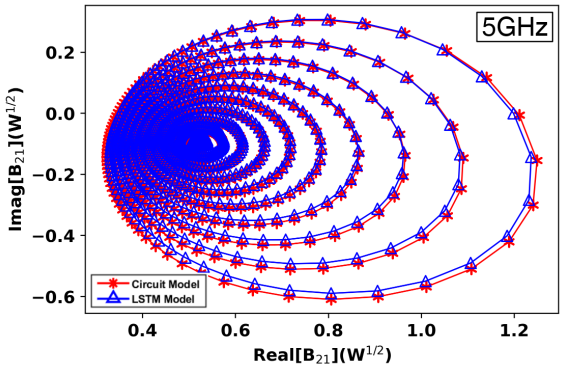


FIGURE 16. Frequency interpolation test of the proposed LSTM model (blue triangles) at 5 GHz, and model extracted with +20 dBm input at 4.5 GHz and 5.5 GHz.

input power levels were +23 and +25 dBm, with the operating frequency set at 1 GHz. Tests were performed when the input power was +24 dBm, and the prediction performance is given in Fig. 15. As can be seen, the model provides excellent predictions. Similarly, the frequency interpolation of the model is also given here. It was extracted when the input power was +20 dBm and the operating frequencies were 4.5 GHz and 5.5 GHz. Tests were performed when the operating frequency was set at 5 GHz with +20 dBm input, and the prediction performance is shown in Fig. 16. From this figure, we can see that the proposed model presents great frequency interpolation capability as well. Detailed information on the average relative errors of the model can be found in Tab. 4. For both input power and frequency interpolation, the results in the table show that the interpolation of the model will decrease accuracy a little bit. However, it will still maintain a high level of accuracy.

B. MEASUREMENT RESULTS

In the measurement validation part, LSTM-based test cases for large-signal models are presented. The test was performed on a test bench with a Keysight N5247B PNA-X combined with a Focus Microwaves load-pull system [37] (see Figure 17).

TABLE 4. Average relative error of LSTM model with different input power and frequency.

Device Status	Model	Average Relative Error (%)
1 GHz	+23 dBm (Direct extracted)	0.14
	+25 dBm (Direct extracted)	0.17
	+24 dBm (Interpolated)	0.39
+20 dBm	4.5 GHz (Direct extracted)	0.19
	5.5 GHz (Direct extracted)	0.21
	5 GHz (Interpolated)	1.92

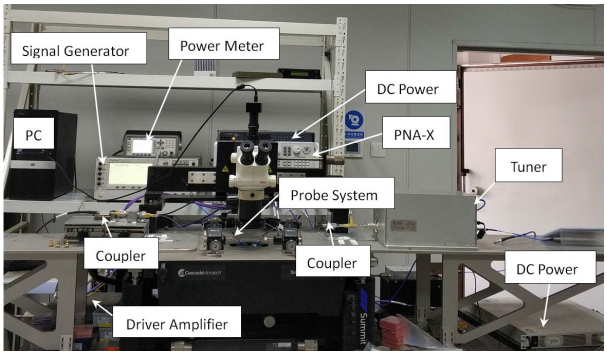


FIGURE 17. Experimental test bench.

In the first measurement validation example, the 10-W GaN HEMT, which is the same device that has been employed in the simulation example, is employed in the test. In the large-signal model validation test, firstly, the input power interpolation capability of the model is tested. In the test, the drain and gate are biased at 28 V and −3 V, respectively. Three different input power levels: +20 dBm, +25 dBm, and +30 dBm at 2 GHz, are used. 108 (9 × 12) sampling points are used at each input power level, for model extraction.

After the model has been trained, it will be available for use to predict the fundamental scattered wave behavior of the DUT. For a given available input power level, 649 different load points are used for testing, and the second- and third-harmonic loads are set to 50. As illustrated in Fig. 18(a), the LSTM model effectively predicts the response, i.e., scattered wave B_{21}, of the device under several excitations, i.e., A_{11} and A_{21}. At these three different input power levels, the newly developed model provides a high level of accuracy throughout the whole Smith chart. Moreover, the capability of predicting time-domain waveforms is also demonstrated in Figures 19 and 20. As can be seen, the proposed LSTM model can accurately model the time-domain behavior, which includes both the fundamental and second harmonic frequencies, where the fundamental and the second harmonic reflection coefficients are fixed at Γ_{21} =0.8/130°, and Γ_{22} =0.95/280°, respectively.

Furthermore, the model's ability to interpolate frequencies was also tested. In the test, the LSTM model was extracted at 1.5 GHz, 2 GHz, and 2.5 GHz. As shown in Fig. 21(a), the model has shown outstanding predictive performance across all three frequencies, with +30 dBm input power. In addition, the extrapolation capability of the LSTM model at different input power levels and operating frequencies is tested. The

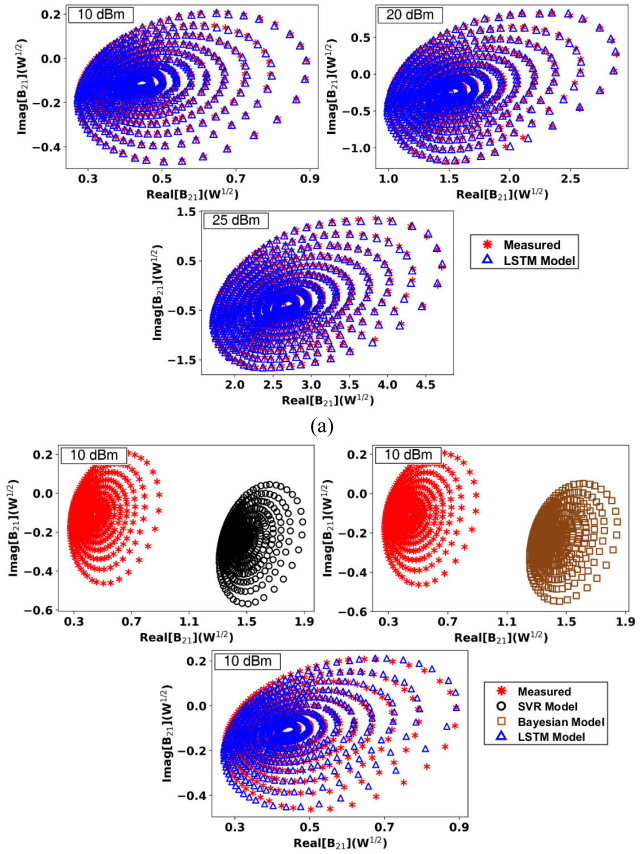


FIGURE 18. (a) Measured data and LSTM model simulations at +10 dBm, +20 dBm, and +25 dBm input with the operating frequency set at 2 GHz. (b) The extrapolation performance of the SVR model, the Bayesian model, and the LSTM model at +10 dBm input by using models extracted with data at +20 dBm input.

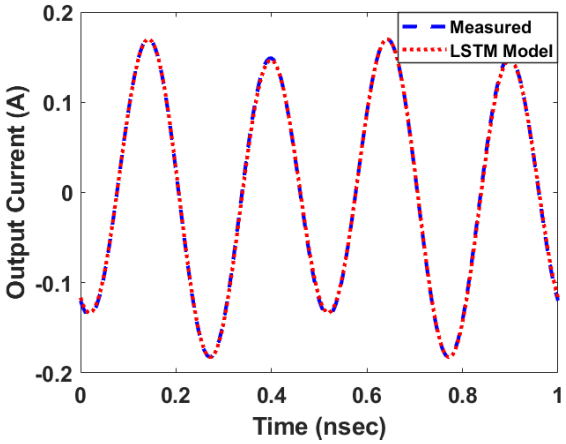


FIGURE 19. Comparison between measurements and LSTM model simulations of the time-domain output current waveform, including fundamental and second harmonic frequencies with Γ21=0.8/130° and Γ22=0.95/280°.

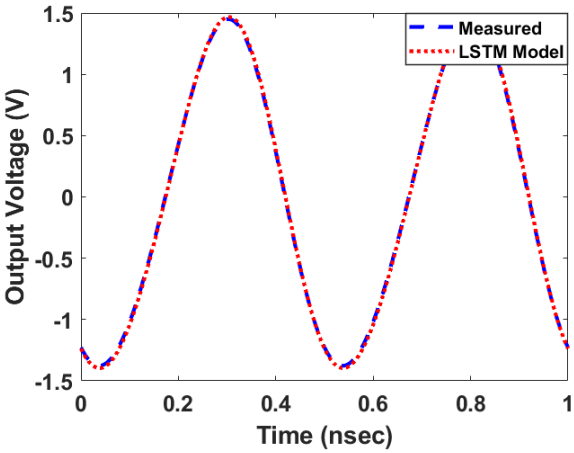


FIGURE 20. Comparison between measurements and LSTM model simulations of the time-domain output voltage waveform, including fundamental and second harmonic frequencies with Γ21=0.8/130° and Γ22=0.95/280°.

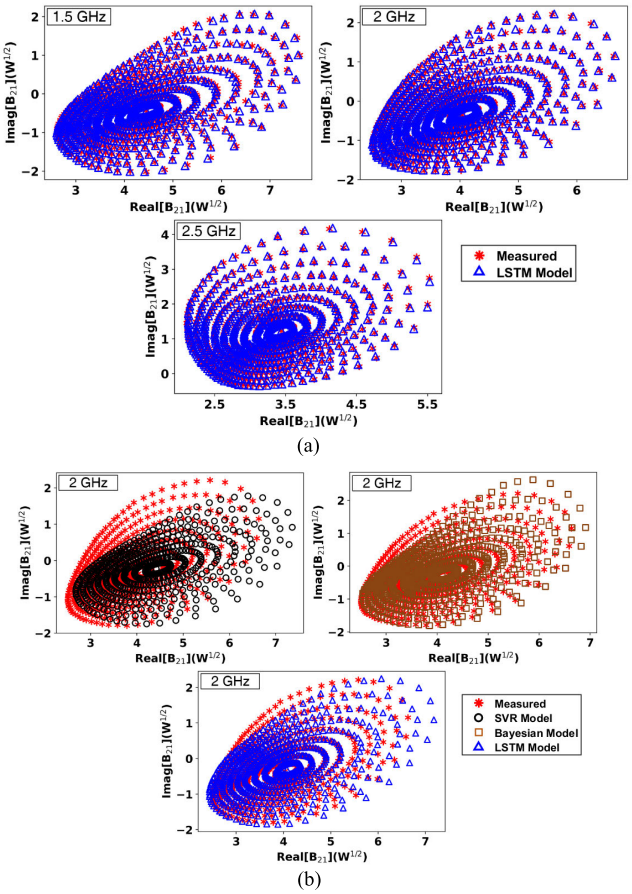


FIGURE 21. (a) Measured data and the LSTM model simulations at 1.5 GHz, 2.0 GHz, and 2.5 GHz with +30 dBm input power. (b) The extrapolation performance of the SVR model, the Bayesian model, and the LSTM model at 2.0 GHz by using models extracted with data at 1.5 GHz.

prediction performance of the SVR model and the Bayesian model is also given as reference cases. In the power extrapolation case, the model was extracted with +20 dBm input and tested when the input power was +10 dBm. Figure 18(b) illustrates the prediction performance achieved. The results show that the LSTM model can accurately match the fundamental scattered waves at load-pull points, although the input power is far from the power used for extraction. On the other hand, the other two models, with their predictions shifting significantly from the measured data, exhibit poor extrapolation performance. For the frequency extrapolation

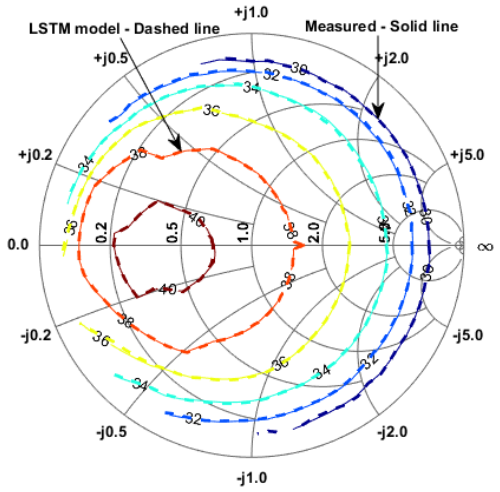


FIGURE 22. Fundamental output power contour with +30 dBm available input power at 2 GHz.

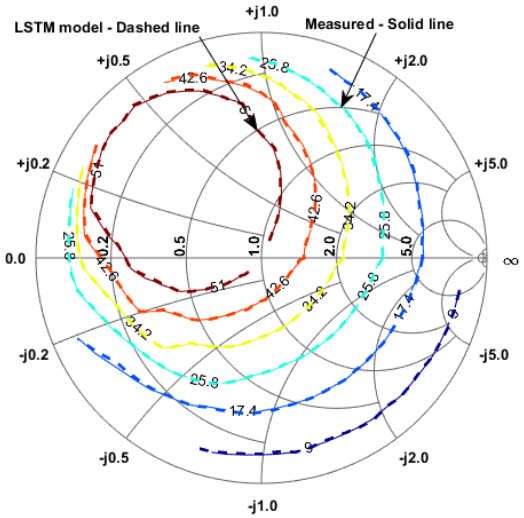
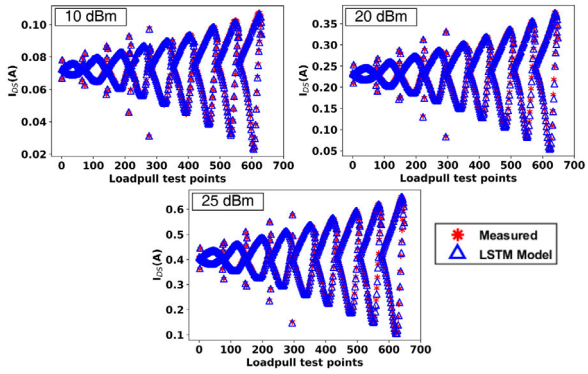


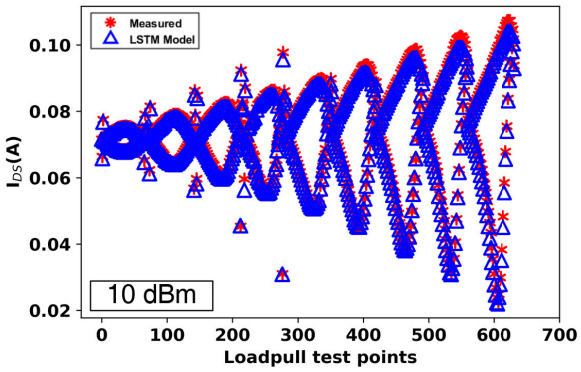
FIGURE 23. Power added efficiency contour with +30 dBm available input power at 2 GHz.

test, it was extracted when the input signal was 1.5 GHz with +30 dBm input power, and the model was tested at 2 GHz. The results of the prediction are shown in Fig. 21(b). The predictions from all three models have deviated from the measured results, though the SVR and Bayesian models have a much more pronounced deviation than the proposed model. The fundamental output power and PAE contour prediction of the proposed model is also presented in Figs. 22 and 23, respectively, when the input power is +30 dBm at 2 GHz. As can be seen from the reported results, the model can accurately predict both the optimal output power and PAE region in the Smith chart for the DUT.

The detailed average relative errors of the model are shown in Tab. 5. The results in the table indicate that the interpolation of input power and frequency has been maintained at a high level. For the input power extrapolation test, the proposed model still gives considerable accuracy when the input power is far from the power point used for extraction. Despite not



(a)



(b)

FIGURE 24. (a) Measured data and the LSTM model simulations of Ids versus load pull test points at +10 dBm, +20 dBm, and +25 dBm input with the operating frequency set at 2 GHz. (b) The extrapolation performance of the LSTM model at +10 dBm input by using the model extracted with data at +20 dBm.

TABLE 5. Average relative error of different models (direct extracted and extrapolated).

Device Status	Model	Average Relative Error (%)
2 GHz	+10 dBm (Interpolated)	0.24
	+20 dBm (Interpolated)	0.31
	+25 dBm (Interpolated)	0.33
	+10 dBm (Extrapolated by LSTM)	1.95
	+10 dBm (Extrapolated by SVR)	229.21
	+10 dBm (Extrapolated by Bayesian)	210.76
+30 dBm	1.5 GHz (Interpolated)	0.32
	2.0 GHz (Interpolated)	0.35
	2.5 GHz (Interpolated)	0.34
	2.0 GHz (Extrapolated by LSTM)	3.11
	2.0 GHz (Extrapolated by SVR)	12.84
	2.0 GHz (Extrapolated by Bayesian)	11.45

performing as well as the power extrapolation test results, the frequency extrapolation results of the model showed reasonable accuracy, less than 4%.

In the last experimental validation example, the performance of the LSTM-based DC model is displayed in Fig. 24(a). Each subplot represents a different input power test case. The model shows acceptable prediction performance in all three test cases. In addition, the extrapolation capability of the proposed LSTM model for DC drain current prediction with different input powers is also validated, as illustrated in Fig. 24(b). The model extracted with

TABLE 6. Average relative error of LSTM DC current model for I_{DS}.

Input Freq	Model	Average Relative Error (%)
2 GHz	+10 dBm (Interpolated)	0.66
	+20 dBm (Interpolated)	0.87
	+25 dBm (Interpolated)	0.96
	+10 dBm (Extrapolated)	1.33

+10 dBm input was used to predict the DC load-pull behavior of the DUT with +20 dBm input. The proposed model makes excellent predictions. Detailed information about the model can be found in the Tab. 6. The results in the table show that the prediction accuracy of the DC drain current decreases slightly with the increase in the input power in the DC condition. However, the average relative error is still less than 1%. Moreover, the DC current extrapolation test results of the model indicate reasonable accuracy, which is less than 2%.

IV. CONCLUSION

This article presents and validates a novel nonlinear behavioral modeling technique for GaN HEMT devices based on LSTM neural networks. Tests and comparisons of different optimization algorithms are conducted in order to identify the optimal optimization algorithm for the model. An example of a 10-watt GaN HEMT device is used to validate the nonlinear behavioral model. As part of this study, several test examples have been selected and the corresponding simulation and measurement results have been presented for RF and DC cases. In all tests, the model has demonstrated a strong capability to extrapolate and interpolate across input power levels and operating frequencies, resulting in the most accurate predictions. Based on comparisons between the proposed model and the two existing models, the proposed model shows superior prediction performance, especially when extrapolation is taken into account.

REFERENCES

[1] A. Jarndal and G. Kompa, "An accurate small-signal model for AlGaN-GaNHEMT suitable for scalable large-signal model construction," *IEEE Microw. Wireless Compon. Lett.*, vol. 16, no. 6, pp. 333–335, Jun. 2006.

[2] D. W. Runton, B. Trabert, J. B. Shealy, and R. Vetury, "History of GaN: High-power RF gallium nitride (GaN) from infancy to manufacturable process and beyond," *IEEE Microw. Mag.*, vol. 14, no. 3, pp. 82–93, May 2013.

[3] L. Dunleavy, C. Baylis, W. Curtice, and R. Connick, "Modeling GaN: Powerful but challenging," *IEEE Microw. Mag.*, vol. 11, no. 6, pp. 82–92, Oct. 2010.

[4] P. M. Cabral, J. C. Pedro, and N. B. Carvalho, "Nonlinear device model of microwave power GaN HEMTs for high power-amplifier design," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 11, pp. 2585–2592, Nov. 2004.

[5] J. B. King and T. J. Brazil, "Nonlinear electrothermal GaN HEMT model applied to high-efficiency power amplifier design," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 444–454, Jan. 2013.

[6] P. Choi, S. Goswami, U. Radhakrishna, D. Khanna, C. C. Boon, H. S. Lee, D. Antoniadis, and L. S. Peh, "A 5.9-GHz fully integrated GaN frontend design with physics-based RF compact model," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 4, pp. 1163–1173, Apr. 2015.

[7] F. Costanzo, A. Piacibello, M. Pirola, P. Colantonio, V. Camarchia, and R. Giofre, "A novel stacked cell layout for high-frequency power applications," *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 6, pp. 597–599, Jun. 2021.

[8] S. Mao, W. Zhang, Y. Yao, X. Yu, H. Tao, F. Guo, C. Ren, T. Chen, B. Zhang, R. Xu, B. Yan, and Y. Xu, "A yield-improvement method for millimeter-wave GaN MMIC power amplifier design based on load—Pull analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 8, pp. 3883–3895, Aug. 2021.

[9] R. S. Pengelly, S. M. Wood, J. W. Milligan, S. T. Sheppard, and W. L. Pribble, "A review of GaN on SiC high electron-mobility power transistors and MMICs," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1764–1783, Jun. 2012.

[10] M. A. Alim, M. M. Ali, A. A. Rezazadeh, and C. Gaquiere, "Temperature dependence of the Taylor series coefficients and intermodulation distortion characteristics of GaN HEMT," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 39, no. 3, pp. 552–559, Mar. 2020.

[11] L. Chen, W. Chen, F. M. Ghannouchi, and Z. Feng, "2-D magnitude-selective affine function-based digital predistortion for concurrent dual-band terminal power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 9, pp. 4209–4222, Sep. 2021.

[12] V. Vadala, A. Raffo, G. Avolio, M. Marchetti, D. M. M.-P. Schreurs, and G. Vannini, "A new dynamic-bias measurement setup for nonlinear transistor model identification," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 1, pp. 218–228, Jan. 2017.

[13] A. Raffo, G. Avolio, V. Vadala, G. Bosi, G. Vannini, and D. Schreurs, "Assessing GaN FET performance degradation in power amplifiers for pulsed radar systems," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 11, pp. 1035–1037, Nov. 2018.

[14] F. M. Barradas, L. C. Nunes, T. R. Cunha, P. M. Lavrador, P. M. Cabral, and J. C. Pedro, "Compensation of long-term memory effects on GaN HEMT-based power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3379–3388, Sep. 2017.

[15] D. D. Mahajan, S. A. Albahrani, R. Sodhi, T. Eguchi, and S. Khandelwal, "Physics-oriented device model for packaged GaN devices," *IEEE Trans. Power Electron.*, vol. 35, no. 6, pp. 6332–6339, Jun. 2020.

[16] A. Jarndal and G. Kompa, "A new small-signal modeling approach applied to GaN devices," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 11, pp. 3440–3448, Nov. 2005.

[17] A. Jarndal, R. Essaadali, and A. B. Kouki, "A reliable model parameter extraction method applied to AlGaN/GaN HEMTs," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 35, no. 2, pp. 211–219, Feb. 2016.

[18] A. S. Hussein and A. H. Jarndal, "Reliable hybrid small-signal modeling of GaN HEMTs based on particle-swarm-optimization," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 9, pp. 1816–1824, Sep. 2018.

[19] D. E. Root, J. Verspecht, J. Horn, and M. Marcu, *X-Parameters*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[20] D. E. Root, J. Verspecht, D. Sharrit, J. Wood, and A. Cognata, "Broadband poly-harmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 11, pp. 3656–3664, Nov. 2005.

[21] J. Cai, J. B. King, B. M. Merrick, and T. J. Brazil, "Padé-approximation-based behavioral modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 12, pp. 4418–4427, Dec. 2013.

[22] J. Cai, J. Liu, J. Su, L. Sun, S. Chen, J. Xia, and J. B. King, "Behavioral model for RF power transistors based on canonical section-wise piecewise linear functions," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 4, pp. 1409–1422, Apr. 2020.

[23] T. M. Martin-Guerrero, A. Santarelli, G. P. Gibiino, P. A. Traverso, C. Camacho-Penalosa, and F. Filicori, "Automatic extraction of measurement-based large-signal FET models by nonlinear function sampling," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 5, pp. 1627–1636, May 2020.

[24] A. Raffo, G. Bosi, V. Vadalá, and G. Vannini, "Behavioral modeling of GaN FETs: A load-line approach," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 1, pp. 73–82, Jan. 2014.

[25] J. Cai, J. B. King, C. Yu, S. Chen, Q. Xie, and L. Sun, "A combined broadband model for GaN HEMTs in admittance domain based on canonical piecewise linear functions," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 12, pp. 5042–5054, Dec. 2020.

[26] M. Geng, J. Cai, and J. B. King, "Modified small-signal behavioral model for GaN HEMTs based on support vector regression," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 31, no. 9, Sep. 2021, Art. no. e22774.

[27] T. R. Turlington, *Behavioral Modeling of Nonlinear RF and Microwave Devices*. Norwood, MA, USA: Artech House, 1999.
[28] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*. Norwood, MA, USA: Artech House, 2014.
[29] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.
[30] J. Verspecht and D. E. Root, "Polyharmonic distortion modeling," *IEEE Microw. Mag.*, vol. 7, no. 3, pp. 44–57, Jun. 2006.
[31] J. Verspecht, D. Gunyan, J. Horn, J. Xu, A. Cognata, and D. E. Root, "Multi-tone, multi-port, and dynamic memory enhancements to PHD nonlinear behavioral models from large-signal measurements and simulations," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2007, pp. 969–972.
[32] Y. Ko, P. Roblin, A. Z.-D. Landa, J. A. Reynoso-Hernandez, D. Nobbe, C. Olson, and F. J. Martinez, "Artificial neural network model of SOS-MOSFETs based on dynamic large-signal measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 3, pp. 491–501, Mar. 2014.
[33] J. Xu, S. Halder, F. Kharabi, J. McMacken, J. Gering, and D. E. Root, "Global dynamic FET model for GaN transistors: DynaFET model validation and comparison to locally tuned models," in *Proc. 83rd ARFTG Microw. Meas. Conf.*, Jun. 2014, pp. 1–6.
[34] H. Kabir, L. Zhang, M. Yu, P. H. Aaen, J. Wood, and Q.-J. Zhang, "Smart modeling of microwave devices," *IEEE Microw. Mag.*, vol. 11, no. 3, pp. 105–118, May 2010.
[35] Z. Marinkovic, G. Crupi, A. Caddemi, V. Markovic, and D. M. M. Schreurs, "A review on the artificial neural network applications for small-signal modeling of microwave FETs," *Int. J. Numer. Model., Electron. Netw., Devices Fields*, vol. 33, no. 3, p. e2668, May 2020.
[36] J. Cai, J. Liu, J. King, and L. Sun, "Machine learning technique based extremely broadband small-signal behavioral model for InP DHBTs," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 30, no. 7, Jul. 2020, Art. no. e22235.
[37] J. Cai, J. B. King, J. Su, C. Yu, S. Chen, L. Sun, H. Wang, and J. Liu, "Bayesian inference-based behavioral modeling technique for GaN HEMTs," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 6, pp. 2291–2301, Jun. 2019.
[38] J. Cai, J. King, C. Yu, J. Liu, and L. Sun, "Support vector regression-based behavioral modeling technique for RF power transistors," *IEEE Microw. Compon. Lett.*, vol. 28, no. 5, pp. 428–430, May 2018.
[39] M. Geng, Z. Zhu, and J. Cai, "Small-signal behavioral model for GaN HEMTs based on long-short term memory networks," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2021, pp. 1–3.
[40] S. Hochreiter and J. Schmidhuber, "Long-short term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[41] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
[42] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.
[43] 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)*, Brive La Gaillarde, France, Jul. 2018, pp. 1–3.
[44] 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.
[45] T.-T.-H. Le, J. Kim, and H. Kim, "An effective intrusion detection classifier using long short-term memory with gradient descent optimization," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Feb. 2017, pp. 1–6.
[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[47] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.
[49] T. Dozat, "Incorporating Nesterov momentum into Adam," in *Proc. ICLR*, Apr. 2016, pp. 1–4.
[50] J. Wang and Z. Cao, "Chinese text sentiment analysis using LSTM network based on L2 and Nadam," in *Proc. IEEE 17th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2017, pp. 1891–1895.
[51] A. Khusro, M. S. Hashmi, and A. Q. Ansari, "Exploring support vector regression for modeling of GaN HEMT," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Kolkata, India, Nov. 2018, pp. 1–3.

MINGQIANG GENG (Student Member, IEEE) was born in Hangzhou, Zhejiang, China, in 1998. He received the B.S. degree from Anhui Polytechnic University, in 2019. He is currently pursuing the master's degree with Hangzhou Dianzi University, Hangzhou. His main research interest includes the modeling of RF power devices with machine-learning algorithms.

GIOVANNI CRUPI (Senior Member, IEEE) received the M.Sc. (cum laude) and Ph.D. degrees in electronic engineering from the University of Messina, Italy, in 2003 and 2006, respectively.

Since 2005, he has been a repeat Visiting Scientist with KU Leuven and IMEC, Leuven, Belgium. He is currently an Associate Professor with the University of Messina, where he teaches/taught bioengineering, microwave electronics, wireless technologies, and optoelectronics. He has authored or coauthored almost 200 publications in international journals and conferences. He has co-edited two books, namely *Microwave De-embedding: From Theory to Applications* (Oxford, U.K.: Elsevier Academic Press, 2013) and *Microwave Wireless Communications: From Transistor to System Level* (Oxford, U.K.: Elsevier Academic Press, 2016). His main research interests include the characterization and modeling of microwave transistors for wireless applications as well as microwave sensors for bioengineering applications.

Dr. Crupi serves as a TPC Member for the IEEE INMMiC, TELSIKS, and ICTA conferences. Since 2012, he has been the Chair of the IEEE MTT-S Graduate Fellowship Program, which is one of the longest continuing education-related activities of the IEEE MTT-S and awards up to 12 annual fellowships (each in the amount of U.S. $6000 plus U.S. $1000 to attend IMS) to encourage future leaders and key technical contributors in their pursuit of careers in RF and microwave engineering. He was the TPC Chair of the IEEE INMMiC, in 2014 and 2015. He is an Associate Editor of IEEE Access and *International Journal of RF and Microwave Computer-Aided Engineering* (Wiley) and a Senior Editor of the *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* (Wiley). He was the Guest Editor of four special issues: two in the IEEE Transactions on Microwave Theory and Techniques, in 2014 and 2016, and two in the *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* (Wiley), in 2014 and 2021.

JIALIN CAI (Senior Member, IEEE) received the B.E. degree in electronic engineering from Zhejiang University, in 2007, the M.E. degree in electronic engineering from Southeast University, in 2010, and the Ph.D. degree in electronic engineering from University College Dublin, in 2015.

From 2015 to 2016, he was a Postdoctoral Researcher with the University of Aveiro, Portugal. He is currently a Professor with the Key Laboratory of RF Circuit and System, Ministry of Education, Hangzhou Dianzi University, Zhejiang, China. His main research interests include active device-, circuit-, and system-level modeling as well as the analysis and design of nonlinear microwave circuits, in particular, the RF power amplifiers. He serves as a TPC member for several IEEE conferences. He is an Associate Editor of the *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields* (Wiley).