Efficient Nonlinear Behavior Modeling Method for Voltage-Variable Capacitors

Sichen Yang, *Student Member, IEEE*, Xuan Chen, *Student Member, IEEE*, Chengming Wang, *Student Member, IEEE*, Quankun Chen, *Student Member, IEEE*, Chenghan Wu, *Student Member, IEEE*, Zhaoyang Feng, *Student Member, IEEE*, Mark D. Butala, *Member, IEEE*, and Er-Ping Li, *Fellow, IEEE*

***Abstract*— A voltage-variable capacitor is an essential part of the matching element in radio frequency (RF) circuits, including the transmitting and receiving channel of mobile devices. However, the nonlinear characteristic of the voltage-variable capacitor produces unwanted harmonics when driven by a large exciting signal, which causes electromagnetic radiation spurious emission (RSE) complications. This article presents an efficient nonlinear capacitance polynomial model method coupling with harmonics test systems to predict the nonlinear behavior of the voltage-variable capacitor. In addition, a simple neural network model is trained to reduce the need to manually obtain the nonlinear parameters. Moreover, for devices exhibiting strong nonlinearity, an improved long short-term memory (LSTM) network is proposed to predict harmonics. Ultimately, the nonlinear parameter prediction for such devices is obtained quickly by implementing our developed method, which provides intuitive guidance in the design phase.**

***Index Terms*— Electromagnetic radiation spurious emission (RSE), neural network, nonlinear characteristics, voltage-variable capacitor.**

I. Introduction

THE nonlinear behavior of radio frequency (RF) circuits/devices plays a vital role in the circuit design with the increasing operating frequency. As a consequence, developing an accurate analysis model for the nonlinear behavior of the RF devices becomes essential [1], [2]. For behavioral model research, there exist two general approaches from a macro perspective: parameterized approaches such as winner models [3], polynomial models [4], and harmonic balance [5] and black-box models such as neural networks, support vector machines, X-parameters [6]–[8], and Volterra series [9], [10], where input and output responses are used to represent the characteristics of devices, modules, and systems such as memory effects and nonlinearities, in the situation when the internal structure is not known. In recent years, the research on the behavior model of RF microwave circuits is mainly devoted to the development of models, extraction of device parameters, improvement of model accuracy, and model practicability, where many achievements have been made.

Over the past decades, a computer-aided design (CAD) approach incorporating artificial neural networks (ANNs) has been recognized as a useful technique for high-frequency electronic circuit modeling [11]–[14] and nonlinear behavioral modeling [15], [16]. Approximation theory, the theoretical basis of neural network technology, that is, a neural network containing at least one hidden layer neuron, can approximate any multidimensional nonlinear continuous function to arbitrary precision. Therefore, the ANN approach has the potential to investigate the nonlinear behavior from measured or simulated input–output data, avoiding the otherwise manual effort of developing equivalent circuit topology. The calculation speed of the neural network model from input to output is faster, which greatly shortens the simulation cycle of the circuit/system with high accuracy. Compared with other approaches such as equivalent circuit and analytical model reduction techniques, ANN models are easily adapted for modeling new devices or new circuits by computerized training algorithms. ANN models can be developed by using the external signals of the original circuit without having to rely on the internal details. This feature is very useful for modeling a new device or circuit when the analytical representation is not available or when the detailed model is too computationally intensive. Compared with other behavior model such as harmonic balance and Volterra series, an ANN can deal with various problems, including weak nonlinearity, strong nonlinearity, and problems with frequency- and time-domain signal data.

These neural-network-based techniques are also more flexible, which can be used to develop more general nonlinear models. Recently, several types of dynamic neural network (DNN) models, e.g., DNNs [16], time-delay neural networks (TDNNs) [17], recurrent neural networks (RNNs) [18], and state-space DNNs (SSDNNs) [19], have been proposed to achieve good accuracy for device nonlinear modeling with nonlinear characteristics such as voltage-variable capacitors.

Voltage-variable capacitors are often used in the matching circuit of mobile device antennas because of their tunable

Manuscript received January 16, 2021; accepted January 30, 2021. Date of publication February 4, 2021; date of current version March 19, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62071424 and Grant 62027805, in part by the Zhejiang Laboratory Foundation of China under Grant 2020KCDAB01, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LD21F010002. Recommended for publication by Associate Editor R. Sharma upon evaluation of reviewers' comments. *(Corresponding author: Er-Ping Li.)*

The authors are with the Department of Information System and Electronic Engineering, ZJU-UIUC Institute, Zhejiang University, Hangzhou 310027, China (e-mail: liep@zju.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCPMT.2021.3057115.

Digital Object Identifier 10.1109/TCPMT.2021.3057115

characteristics. The high-permittivity thin-film dielectrics in the capacitor may exhibit strong field dependence in the dielectric constant, which will result in nonlinear problems in the RF circuit [20], [21]. Moreover, the unwanted harmonics generated due to a large excited signal will lead to radiated spurious emission (RSE) [22], [23]. Once the RSE exceeds the limits, the device will not function well. Therefore, modeling the electromagnetic radiation and the nonlinear performance of the voltage-variable capacitor is vitally important and will allow us to better predict various characteristics of the device under the design phase.

This article presents a nonlinear polynomial modeling method for the voltage-variable capacitor mentioned above. In this model, the nonlinear parameters of the capacitor are obtained using the commercial simulation simulator advanced design system (ADS) optimizer combined with the nonlinear test system. In order to improve efficiency, a multilayer ANN model has trained to get nonlinear parameters quickly.

In the case of strong nonlinearity such as mixers [24], the order of magnitude of the harmonics generated by the signal after passing through the device is close to that of the fundamental wave, and the values between the harmonic waves as well as to the fundamental wave will affect each other. Therefore, for a complex problem with strong nonlinear behavior, an improved long short-term memory (LSTM) network has been proposed to learn the nonlinear behavior since ordinary neural networks cannot accurately capture such strong nonlinear relationships.

In this article, Section II presents the nonlinear test system and the deembed method. Section III presents the mechanism for nonlinear characterization of the voltage-variable capacitor and ADS modeling method and results of the voltage-variable capacitor. Section IV provides the multilayer neural network model trained by data described in Sections II and III, demonstrating our faster and more flexible modeling method. Section V provides the improved LSTM network trained by simulation data from ADS. Our concluding remarks are given in Section VI.

II. Test System and Process

Fig. 1 shows the nonlinear test flow using automatic-fixture-removal (AFR) calibration, where the phase of the fundamental wave and harmonics are not concerned, and only magnitude using a spectrum analyzer is measured.

Fig. 2 shows the harmonics test block diagram with the main test instruments, including a signal generator, spectrum analyzer, and power meter.

When measuring high-power devices, a power amplifier and the passive components need to be added to the test block diagram to meet the requirements and safety considerations for high-power testing. The input channel is used to filter the harmonics of the input signal and isolate the reflected waves and the diplexer separates the fundamental and harmonic waves. The frequency points interested is the Global System for Mobile Communications (GSM) frequency at 900 MHz and its vicinity. Two different lengths of microstrip lines as different fixtures are measured to validate the consistency of the measurement results for the voltage-variable capacitor.

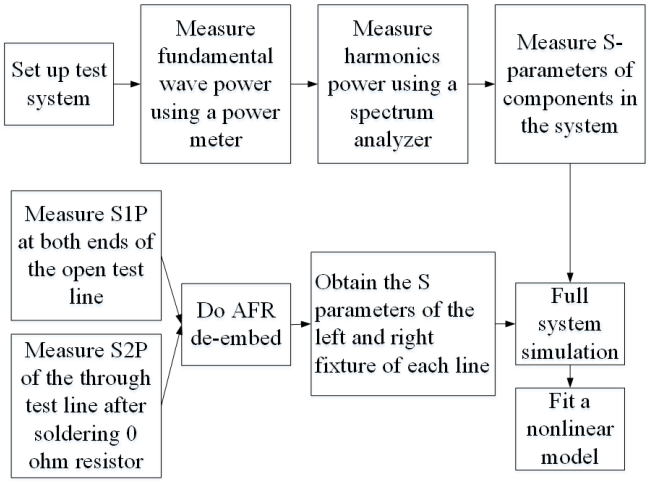

Fig. 1. Nonlinear test flow using AFR calibration.

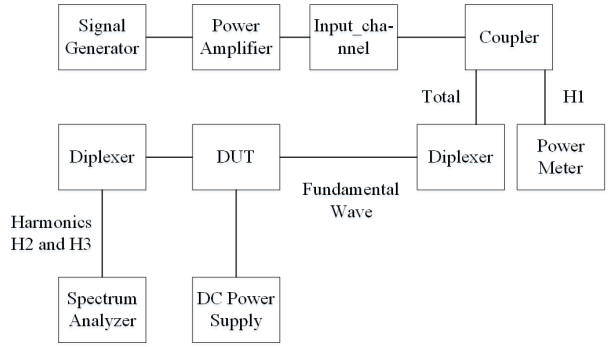

Fig. 2. Schematic of the test system.

The power magnitude of the fundamental wave ($H1$) input to the capacitor and the power values of the second- and third-harmonic waves ($H2$ and $H3$) of the capacitor output are obtained from the power meter and spectrum analyzer, respectively. Measuring S-parameters of various components and devices of the test system is necessary for the nonlinear parameters fitting stage. It is worth mentioning that as the dc voltage changes, its capacitance value also changes, and the nonlinear characteristics are naturally different. Therefore, a dc power supply is needed to control the dc input from 1 to 18 V.

For deembedding, two common methods thru-reflect-line (TRL) [25] and AFR [26] are used. After a comparison, the AFR method is selected since its accuracy is generally higher, possibly because TRL only tests the S-parameters of a fixed length fixture such as 5 cm. For other fixtures with different lengths, it would have to employ electromagnetic simulation software such as CST and HFSS to obtain the results. To measure the S-parameters of each length fixture, the corresponding number and length of TRL calibration lines have to be manufactured. For AFR, open and through conditions of each fixture are measured to get the S-parameters for different length fixtures in AFR, as shown in Fig. 1. First, the test line is open, and one-port S-parameter at both ends of the test line is measured. Then, a 0-Ω resistance is soldered to

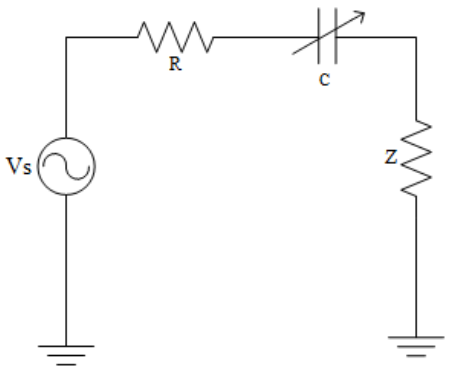

Fig. 3. Simple circuit contains a nonlinear capacitance.

connect two fixtures to measure two-port S-parameter of the through test line. Then, AFR deembed is done to obtain the S-parameters of the left and right fixtures of each line.

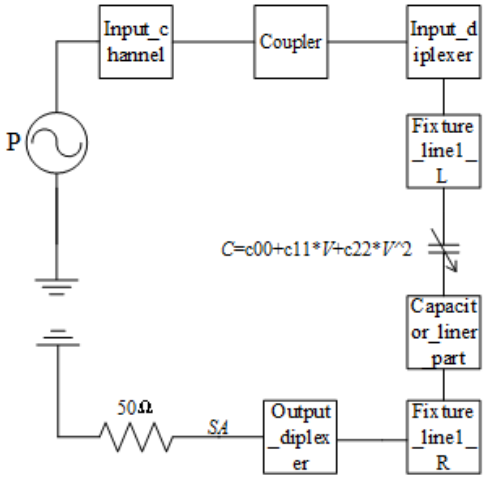

Fig. 4. ADS block diagram of AFR-based nonlinear capacitance fitting model.

III. Nonlinear Capacitance Model

A. Nonlinear Capacitance Polynomial Model Expression

Referring to the closed-form expression for the capacitance–voltage nonlinearity [20], the nonlinearity model of a voltage-variable capacitor uses power series and the C–V characteristic is expressed as

$$C(V) = \frac{C_{\max}}{2\cosh\left[\frac{2}{3}\sinh^{-1}\left(\frac{2V}{V_2}\right)\right] - 1} \tag{1}$$

where V is the voltage across the capacitor, $C_{\max}$ is the peak capacitance at zero applied field, and V_2 is the voltage at which the capacitance is half of the max capacitance.

Its Taylor series expansion for the third order of $C(V)$ has a constant and the second-order terms expressed as

$$C(V) = C_{\max} - \frac{16C_{\max}V^2}{9V_2^2}. \tag{2}$$

This voltage-variable capacitor is simulated as a nonlinear capacitance contains a constant term and second-order term. Refer to (2) and circuit theory, the test result of the output signal of the voltage-variable capacitor should contain only third harmonic.

Taking Fig. 3 as an example, the circuit contains a single-tone signal source V_s, source impedance R, nonlinear capacitance C, and load impedance Z. The relationship is written as

$$C\frac{dV_C}{dt}(R+Z) + V_C = V_s \tag{3}$$

where

$$C(V_C) = C_0 + C_2V_C^2 \tag{4}$$

where C_0 and C_2 are arbitrary constant. Refer to the harmonic balance method, the circuit current and the voltage across the components have fundamental wave term and third-harmonic wave but no second-harmonic wave. However, it is observed that the actual test result shows a second-harmonic wave, that is, the voltage at both ends of the capacitor is not only dc for tuning the capacitance but also large ac voltage for excitation at a test case. Therefore, when substituting $V\_ac + V\_bias$ for V in (1), where V_ac and V_bias are the ac and dc voltage, respectively, at both ends of the capacitor. The capacitance C is expressed with a constant term, first-order term, and second-order term by the Tylor expansion

$$C(V\_ac) = C_0 + C_1V\_ac + C_2V\_ac^2 \tag{5}$$

where C_1 and C_2 are related to V_bias. The above analysis is validated via the following method. When $V\_bias = 0$, it means that the dc power supply is zero, and no second-order harmonic is measured by the spectrum analyzer. When V_bias is nonzero, no matter how small, a second-order harmonic is measurable on the spectrum analyzer. Therefore, the voltage-variable capacitor can be modeled as a nonlinear capacitance polynomial model, which contains constant term C_0, first term C_1, and second term C_2 as shown in (5).

B. Fitting the Nonlinear Capacitance Parameters

Fig. 4 shows the complete test link S-parameters into ADS to express the complete characteristics of the test system. These boxes represent S-parameters, which can adequately represent the linear characteristics of the entire system. The characteristic of the nonlinear part is provided by the nonlinear capacitance in the diagram. The coefficients $c00$, $c11$, and $c22$ of the nonlinear capacitance represent C_0, C_1, and C_2 in (5). SA in the diagram means the power magnitude of the spectrum analyzer from the test. The power source P is the power provided by the signal generator and power amplifier. Through harmonic balance simulation and fitting optimization components, the input power is fitted first in light of the first-order large capacitance C_0 provided by the manufacturer's datasheet and the fundamental wave power magnitude $H1$ measured with the power meter. The second- and third-order capacitances, that is, nonlinear parameters, C_1 and C_2, are fitted using the measured input power and the second- and third-order harmonics measured from the spectrum analyzer.

TABLE I
MODELED PARAMETERS

dc (V)	Capacitor1 (pF) (C_{max} =3.14 pF, V_2 =3.6 V)				
	Line1		Line2		C_2 Error
	C_1	C_2	C_1	C_2	
6	8.77E-5	4.06E-5	1.61E-5	4.05E-5	0.23%
9	8.47E-5	1.72E-5	8.68E-6	1.76E-5	2.45%
12	5.93E-5	8.13E-6	4.94E-6	8.44E-6	3.64%
16	4.19E-5	4.19E-6	3.76E-6	4.47E-6	6.34%

dc (V)	Capacitor2 (pF) (C_{max} =11 pF, V_2 =3 V)				
	Line1		Line2		C_2 Error
	C_1	C_2	C_1	C_2	
5	9.58E-5	1.15E-4	6.02E-5	1.13E-4	1.16%
7	6.69E-5	5.55E-5	4.29E-5	5.53E-5	0.38%
11	4.06E-5	1.88E-5	2.73E-5	1.92E-5	2.14%
16	2.71E-5	6.74E-6	1.83E-5	6.92E-6	2.68%

dc (V)	Capacitor3 (pF) (C_{max} =5.06 pF, V_2 =4.1 V)				
	Line1		Line2		C_2 Error
	C_1	C_2	C_1	C_2	
4	4.19E-6	5.78E-6	6.97E-6	6.21E-6	6.92%
8	2.73E-6	1.66E-6	3.49E-6	1.76E-6	5.6%
11	2.09E-6	6.55E-7	1.56E-6	6.69E-7	2.01%
15	1.74E-6	2.73E-7	1.48E-6	2.79E-7	2.15%

As shown in Table I, three commonly used voltage-variable capacitor patterns with different values of $C_{\max}$ and V_2 in (1) are modeled at the fundamental frequency at 900 MHz. Two different lengths of microstrip lines as different fixtures named Line1 and Line2 are measured to validate the consistency of the measurement results.

Furthermore, the order of C_2 is round 10^{-7}–10^{-4} pF, which looks small. However, due to the large input signal, the second and third harmonics induced will reach almost -40 to -25 dBm as shown in Fig. 5, which is a catastrophic RSE problem. The consistency of C_2 is very good, and all errors are almost below 5% (for the full range of test data from 1 to 18 V). The error may be caused by several factors as follows.

1) *Differences in Individual Voltage-Variable Capacitors*: There are some differences between individual voltage-variable capacitors and the nonlinear capacitance parameters that are small and easily affected by small differences.
2) *Small Deviations in Solder Joints*: The AFR method is obtained by welding a 0-Ω resistance through line for deembedding. Then, the resistance is removed to solder device under test (DUT) to measure the harmonics. This will naturally produce some errors.
3) *Input Power Differences*: The power is obtained by ADS fitting, and some inevitable deviations will occur in the fitting process.

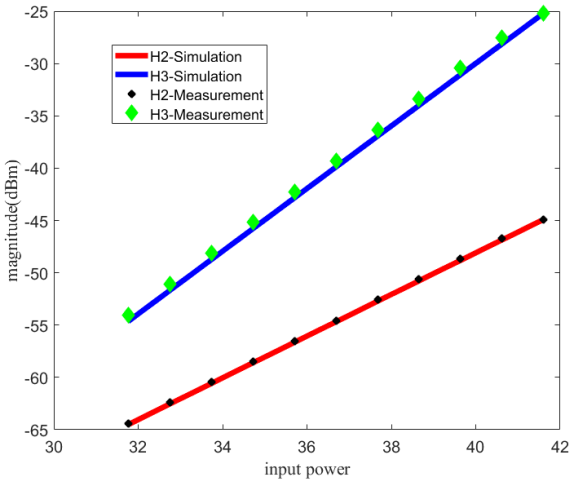

Fig. 5. Set of comparison between simulation data and measurement data of $H2$ and $H3$.

The error of C_1 value is a larger since Line2 is too long which cause higher loss. The measured $H2$ of Line2 in the spectrum analyzer is lower than -75 dBm, which is almost as strong as background noise. As a result, the modeling parameters of Line1 are used for analysis.

Fig. 5 shows one set of comparison of simulation data using the value of fitted nonlinear parameters in Table I from ADS simulation as shown in Fig. 4 and experimental data as shown in Fig. 2 of $H2$ and $H3$, which exhibits the great consistency between simulation and measurement. To a certain extent proved the effectiveness of nonlinear parameters.

IV. MULTILAYER NEURAL NETWORK MODEL

As described in Section III, considerable effort is required to manually manipulate ADS to fit and optimize the nonlinear model. If more studies are required at different dc voltage values, more fixtures, and more fundamental frequency points, so more effort is needed. ANNs are nonlinear networks that are used in representing RF and microwave component behaviors because of their nonlinear modeling and parallel processing capabilities. They can provide a robust representation of the nonlinear mapping between input parameters and the nonlinear parameters.

A. Concept of a Multilayer Neural Network

Here, let n and m represent the number of input and output neurons of a neural network. Let $\boldsymbol{x}$ be an n-vector containing

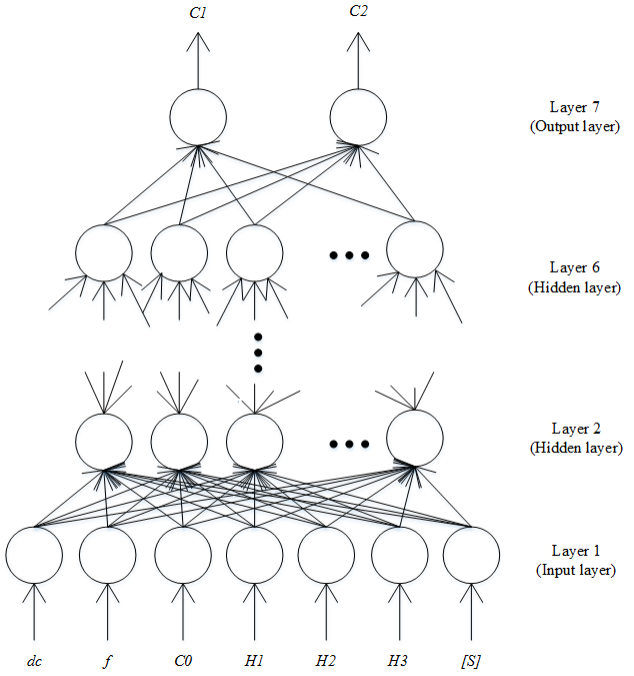

Fig. 6. Multilayer neural network.

the external inputs to the neural network, $\boldsymbol{y}$ be an m-vector containing the outputs from the output neurons, and $\boldsymbol{\omega}$ be a vector containing all the weight parameters representing various interconnections in the neural network.

As described in Section III, the frequency and the linear and nonlinear characteristics of the capacitor itself as well as the input power affect the output fundamental wave and harmonic waves. Conversely, the nonlinear parameter is determined by the frequency, the linear characteristics of the capacitor itself, and the output fundamental wave and harmonic waves. The linear characteristics of the capacitor contain dc voltage across the capacitor and S-parameter of the capacitor. The input power is determined by the output fundamental wave and the first-order large capacitance C_0. Inputs and outputs of the corresponding neural model are given by

$$\boldsymbol{x} = [\text{dc}, f, C_0, H1, H2, H3, [S]]^{\boldsymbol{T}} \tag{6}$$

$$\boldsymbol{y} = [C_1, C_2]^{\boldsymbol{T}} \tag{7}$$

where the relevant inputs are dc voltage dc, fundamental frequency f, first-order large capacitance C_0, S-parameter of the linear part of the capacitor and S-parameter of the fixture [S], fundamental wave magnitude $H1$, and the second- and third-order harmonics magnitudes $H2$ and $H3$. The outputs are second- and third-order small capacitances, C_1 and C_2, that are also the sought after nonlinear parameters. The relationship between the input and the output is nonlinear as

$$\boldsymbol{y} = F(\boldsymbol{x}, \boldsymbol{\omega}). \tag{8}$$

Refer to the approximation theory, the function $\boldsymbol{F}$ in (8) can be solved by a multilayer neural network shown in Fig. 6. The circles represent neurons, the neurons in the first layer are input data, the neurons in the seventh layer are output data, the arrows in the figure represent the weights $\boldsymbol{\omega}$, and the neurons in the middle are hidden layers that perform a nonlinear function. Original $\boldsymbol{x}$ and $\boldsymbol{y}$ data are used to adjust neural network weights $\boldsymbol{\omega}$ such that the neural model outputs best match the training data outputs [27].

For the neural network to be an accurate model of the problem to be learned, a suitable number of hidden neurons are needed. An automatic cycle program is used to add the number of hidden layers until the accuracy changes by less than 0.1%. Finally, a seven-layer neural network is defined, including an input layer, an output layer, and five hidden layers.

TABLE II
MEASURED DATA FOR NEURAL NETWORK TRAINING

Input parameters	***range***
dc	1~18 V
f	0.85, 0.87, 0.88, 0.9, 0.93 and 0.95 GHz
C_0	0.674 pF to 11 pF (determined by dc, C_{max} and V_2 in Eq. (1))
H1	-5 dBm to 10 dBm (measured by spectrum analyzer)
H2	-68 dBm to -35 dBm (measured by spectrum analyzer)
H3	-70 dBm to -29 dBm (measured by spectrum analyzer)
[S]	measured by Vector Network Analyzer

B. Training and Results

Rectified linear unit (ReLU) [28] is used for the activation function, and Adam optimizer [29] is used for the optimizer with the neural-network training error

$$\boldsymbol{E} = \frac{1}{2}\sum\sum|\boldsymbol{y} - \boldsymbol{prediction}|^2 \tag{9}$$

where $\boldsymbol{prediction}$ is the output of the neural network and $\boldsymbol{y}$ is the training data C_1 and C_2 [see (7)]. The machine learning process executes a gradient descent based on the training error and adjusts the coefficients of the neural network to improve the agreement between the output and the label [30].

Two hundred and ninety-nine sets of data from measurement are considered in which 210 for training and the remainder for testing. Also, this has extra 71 sets of data separated from the 299 sets for validation of model extrapolation capabilities. The data range is shown in Table II. $H1$, $H2$, and $H3$ are determined by input power and the characteristics of capacitor and measured by a spectrum analyzer. For training data, the fundamental frequencies considered were 0.85, 0.88, 0.9, and 0.93 GHz, and the dc voltage range is 1–17 V excluding 7 V. For validation data, consider the cases where the fundamental frequency is 0.87 and 0.95 GHz and the dc voltage is 7 and 18 V.

As the results shown in Table III, the interpolation result is good, but extrapolation still has a certain degree of error. The extrapolation for dc voltage is better than for frequency because the number of samples for the dc voltage is larger

TABLE III
ERROR OF PREDICTED NONLINEAR PARAMETERS FROM THE NEURAL NETWORK (NN) MODEL

	Relative Error
Interpolation	4.59%
Extrapolation (f=0.87 GHz)	10.96%
Extrapolation (f=0.95 GHz)	19.36%
Extrapolation (dc=7 V)	6.12%
Extrapolation (dc=18 V)	8.41%

and the regularity of the dc voltage is greater than that of frequency. This result is consistent with the measurement data. When the dc voltage increases, the nonlinear coefficient decreases. When the frequency data change, the nonlinear coefficient changes irregularly. Two values are used with the largest error to substitute into the schematic in Fig. 4. The simulated harmonics differed by only 2%.

The benefits of training this neural network are as follows: reducing the manual operations and increasing efficiency and speed has the ability of interpolation and extrapolation, and similar results can be obtained through neural network without measurement. Moreover, the network can be updated at any time according to the change of the circuit. For example, the input and output of the neural network can be changed when the load impedance changes or the DUT model needs to be changed like nonlinear resistance.

V. IMPROVED LSTM NETWORK

An RNN is a special type of neural network that has the ability to learn and then express the behavior of dynamic systems. RNN has a time correlation in the time domain. The current output is not only related to the current input but also related to the previous input. RNN neural network is successfully employed [31] to establish a behavioral model of a nonlinear microwave circuit, where the training data used time-domain input and output data.

When the data are too large, the problem of gradient disappearance often occurs when using an RNN [32]. Normally, after two or three steps of error propagation, the attenuation will be very strong. LSTM [33] is a special kind of RNN developed mainly to solve the problem of gradient disappearance and gradient explosion during long sequence training.

A. Results for Weak Nonlinear Data

As shown in Section IV, the order of magnitude of nonlinear capacitance parameters C_1 and C_2 is round $10^{-7} \sim 10^{-4}$ pF. The order of magnitude of first-order large capacitance C_0 is approximately 1 pF. This means that the order of magnitude of fundamental wave $H1$ and second- and third-harmonic waves $H2$ and $H3$ differs greatly. The weak nonlinearity for harmonics and fundamental wave has minimum influence on each other. The signal of weak nonlinear case in time domain is smooth.

TABLE IV
ERROR OF PREDICTED SPECTRUM OF WEAK NONLINEAR DATA

	Relative Error
RNN	1.39%
LSTM	1.28%

For weak nonlinearity case, one of the nonlinear capacitance models in Table I was used and measured 640 sets of data in which the excitation signal change from 0.85 to 1-GHz step 0.01 GHz for frequency and from 1 to 40 V step 1 V for magnitude and 500 sets of data for training and 140 sets for testing. The data are based on the input signal and output signal of the capacitor in time domain

$$\boldsymbol{x} = \left[x^1, x^2, \ldots, x^N\right] \tag{10}$$

$$\boldsymbol{y} = \left[y^1, y^2, \ldots, y^N\right] \tag{11}$$

where N is the time steps and x and y are the input signal and output signal, respectively, in the time domain obtained from the inverse discrete Fourier transform of the voltage in the frequency domain. To be clear, our original data are represented in the frequency domain. while the voltage magnitude of the excitation signal and the output signal is in the time domain. The first-, second-, and third-order spectrum are tested due to equipment testing bandwidth issues, and the signal spectrum after the third order is very small and can be ignored. RNN and LSTM are used for training. After training and testing, the signal spectrum can be obtained from the discrete Fourier transform of predicted time-domain output results. As the results shown in Table IV, RNN and LSTM have similar good capabilities in processing weak nonlinear data since the relationship between the input and output signal is simple and has low dependence between waves of various orders.

B. Improved LSTM Network for Strong Nonlinear Data

For strong nonlinearity problem, the magnitude order of nonlinear capacitance parameters C_1 and C_2 approaches to C_0 which means that the order of magnitude of second- and third-harmonic waves $H2$ and $H3$ is close to $H1$, which means that between the harmonics self and to the fundamental wave will affect each other. The same method as Section IV is used to deal with strongly nonlinear data; 640 sets of data from the simulation are used, of which 500 are used for training and the remainder for testing. The fundamental frequency range is from 0.85 to 1 GHz, and the magnitude is from 1 to 40 V. As the results shown in Table IV and Fig. 7, RNN and LSTM have bad capabilities in processing strong nonlinear data since the relationship between the input and the output signal is complex and has high dependence between waves of various orders. LSTM has better results because it can handle data with strong dependence in time domain [33].

Therefore, refer to the strong nonlinear characteristics, an improved LSTM structure is proposed, as shown in Fig. 7. The input and output are sampled at a specific time steps in

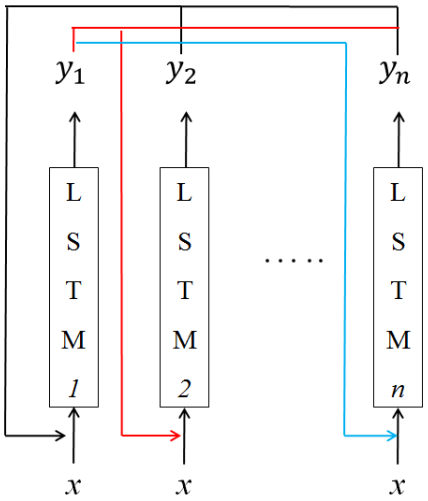

Fig. 7. Improved LSTM network.

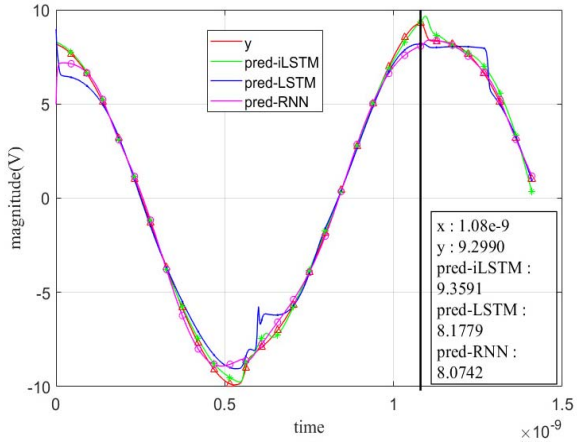

Fig. 8. Comparison of representative predicted and original output.

time domain

$$\boldsymbol{x} = \left[x^1, x^2, \ldots, x^N\right] \tag{12}$$

$$\boldsymbol{y_1} = \left[y_1^1, y_1^2, \ldots, y_1^N\right] \tag{13}$$

$$\boldsymbol{y_2} = \left[y_2^1, y_2^2, \ldots, y_2^N\right] \tag{14}$$

$$\boldsymbol{y_n} = \left[y_n^1, y_n^2, \ldots, y_n^N\right] \tag{15}$$

where N is the time step, n is the highest order nonnegligible term of the output signal, and $\boldsymbol{x}$ is the input signal in the time domain obtained from the inverse discrete Fourier transform of the voltage in the frequency domain. Similarly, $\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_n}$ is the output signal of the nonlinear circuit in the time domain obtained from the inverse Fourier transform of the fundamental wave and harmonics magnitudes in the frequency domain, respectively. The fundamental wave and harmonics are separately fitted because it can better fit the effect of their mutual influence, and such curves are smoother and have higher accuracy. First, the LSTM network is trained for fundamental wave and harmonic waves of each order, and n LSTM networks for n waves can be obtained. Once they all have better accuracy, n LSTM networks would be combined together for training because each wave affects each other. After training and obtaining the harmonics of each order, add them together and compare them with the original time-domain data. The quality of the model can be represented by the mean squared error:

$$\text{MSE} = \frac{1}{N}\sum_{i=0}^{N-1}\left(y^i - \text{pred\_}y^i\right)^2 \tag{16}$$

where y^i and pred_y^i are the original and predicted i step of output signal in time domain, respectively. In this article, TensorFlow [34] is used to train the network. Then, the maximum value of the input signal is used to normalize the original signal since that the maximum value of the fundamental wave and harmonics of the input and output signals are almost equal through simulation. After processing, all the data were found to fall in the range of -1 to 1. The prediction is then multiplied by the maximum value to restore the output signal to its original units. Data preprocessing can make the neural network perform better. When training the neural network, the process execution time is determined by the learning rate. When the learning rate is low, a large number of iterations are required to reach convergence. When the learning rate is high, the search process can be unstable and the parameters will wander about their respective optimal values. Exponential decay for learning rate attenuation is introduced in the training process to obtain good results. As for the number of hidden layer nodes, we refer to some empirical formulas [35] for guidance and still need to adjust experimentally to obtain a good result.

TABLE V
ERROR OF PREDICTED SPECTRUM OF STRONG NONLINEAR DATA

	Relative Error
RNN	32.73%
LSTM	24.86%
Improved_LSTM	5.21%

Fig. 8 shows a random test set of predicted and actual output curve for RNN, LSTM, and improved LSTM network. As the results shown in Table V, the improved LSTM method has better ability to fit strong nonlinear data since it separated waves of each order and considered the effect of each other.

VI. CONCLUSION

In this article, modeling the nonlinear second and third harmonics of a voltage-variable capacitor is successfully developed under the excitation of a large-signal single tone for the proposed test system. The second- and third-order small-capacitance model is optimized and fit using ADS by combining the power series and the test system. The consistency of the third-order small capacitance is verified under different fixtures. In order to simplify manual operation, a multilayer neural network is trained to obtain nonlinear parameters. The interpolation and extrapolation ability is good. The method can quickly and accurately model the voltage-variable capacitor. Moreover, the proposed LSTM network is able to handle the data given in both time and frequency domains, as well

as the problems with strong or weak nonlinearities. This proposed improved LSTM network can deal with all kinds of problems comprehensively, and it has unparalleled advantages in accuracy and speed. The efficient neural network nonlinear macromodel and improved LSTM network for nonlinear performance of voltage-variable capacitors proposed by this article has a great significance in the measuring and modeling of nonlinear devices in the RF frequency.

REFERENCES

[1] C. Baylis, R. Marks, J. Martin, H. Miller, and M. Moldovan, "Going nonlinear," *IEEE Microw. Mag.*, vol. 12, no. 2, pp. 55–64, Apr. 2011.

[2] J. Verspecht, F. Verbeyst, and M. V. Bossche, "Network analysis beyond S-parameters: Characterizing and modeling component behaviour under modulated large-signal operating conditions," in *Proc. 30th Eur. Microw. Conf.*, Oct. 2000, pp. 1–4.

[3] W. Liu, W. Na, L. Zhu, J. Ma, and Q.-J. Zhang, "A Wiener-type dynamic neural network approach to the modeling of nonlinear microwave devices," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 6, pp. 2043–2062, Jun. 2017.

[4] S. A. Mass, *Nonlinear Microwave Circuits*. Norwood, MA, USA: Artech House, 1988.

[5] A. Tesi, E. H. Abed, R. Genesio, and H. O. Wang, "Harmonic balance analysis of period-doubling bifurcations with implications for control of nonlinear dynamics," *Automatica*, vol. 32, no. 9, pp. 1255–1271, Sep. 1996.

[6] J. Verspecht, D. Schreurs, A. Barel, and B. Nauwelaers, "Black box modelling of hard nonlinear behavior in the frequency domain," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 1996, pp. 1735–1738.

[7] Verspecht and D. E. Root, "Polyharmonic distortion modeling," *IEEE Microw. Mag.*, vol. 7, no. 3, pp. 44–57, Jun. 2006.

[8] G. Sun, Y. Xu, and A. Liang, "The study of nonlinear scattering functions and X-parameters," in *Proc. Int. Conf. Microw. Millim. Wave Technol.*, May 2010, pp. 1086–1089.

[9] A. Zhu, M. Wren, and T. J. Brazil, "An efficient volterra-based behavioral model for wideband RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 2, pp. 787–790, 2003.

[10] A. Carini, E. Mumolo, and G. L. Sicuranza, "V-vector algebra and its application to Volterra-adaptive filtering," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 46, no. 5, pp. 585–598, May 1999.

[11] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Norwood, MA, USA: Artech House, 2000.

[12] J. E. Rayas-Sanchez, "EM-based optimization of microwave circuits using artificial neural networks: The state-of-the-art," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 1, pp. 420–435, Jan. 2004.

[13] P. Burrascano, S. Fiori, and M. Mongiardo, "A review of artificial neural networks applications in microwave computer-aided design (Invited article)," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 9, no. 3, pp. 158–174, May 1999.

[14] H. Ma, E.-P. Li, A. C. Cangellaris, and X. Chen, "Support vector regression-based active subspace (SVR-AS) modeling of high-speed links for fast and accurate sensitivity analysis," *IEEE Access*, vol. 8, pp. 74339–74348, 2020.

[15] L. Zhang, J. Xu, M. C. E. Yagoub, R. Ding, and Q.-J. Zhang, "Efficient analytical formulation and sensitivity analysis of neuro-space mapping for nonlinear microwave device modeling," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 9, pp. 2752–2767, Sep. 2005.

[16] J. Xu, M. C. E. Yagoub, R. Ding, and Q. J. Zhang, "Neural based dynamic modeling of nonlinear microwave circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 12, pp. 2769–2780, Dec. 2002.

[17] 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.

[18] Y. Cao and Q.-J. Zhang, "A new training approach for robust recurrent neural-network modeling of nonlinear circuits," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 6, pp. 1539–1553, Jun. 2009.

[19] Y. Cao, R. Ding, and Q.-J. Zhang, "State-space dynamic neural network technique for high-speed IC applications: Modeling and stability analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 6, pp. 2398–2409, Jun. 2006.

[20] D. R. Chase, L.-Y. Chen, and R. A. York, "Modeling the capacitive nonlinearity in thin-film BST varactors," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 10, pp. 3215–3220, Oct. 2005.

[21] J. D. Baniecki *et al.*, "Hydrogen induced tunnel emission in Pt/$(Ba_xSr_{1-x})Ti_{1+y}O_{3+z}$/Pt thin film capacitors," *J. Appl. Phys.*, vol. 89, no. 5, pp. 2873–2885, Mar. 2001.

[22] A. N. Nikitenko, I. V. Ruzhentsev, and V. N. Zin'kovsky, "Level analysis and control of electromagnetic radiation from microwave oscillators," in *Conf. Precis. Electromagn. Meas. Dig.*, Washington, DC, USA, Jul. 1998, pp. 600–601.

[23] B. Audone and R. Colombo, "Measurement of radiated spurious emissions with the substitution and field strength test methods," *Proc. Int. Symp. Electromagn. Compat. EMC Eur., Wroclaw*, Sep. 2016, pp. 353–357.

[24] I. D. Robertson, "Mixers," in *Proc. IEE Colloq., Monolithic Microw. Integr. Circuits (MMIC's)*, 1995, pp. 71–76.

[25] G. F. Engen and C. A. Hoer, "Thru-reflect-line: An improved technique for calibrating the dual six-port automatic network analyzer," *IEEE Trans. Microw. Theory Techn.*, vol. 27, no. 12, pp. 987–993, Dec. 1979.

[26] E. Vandamme, D. M. M.-P. Schreurs, and G. Van Dinther, "Improved three-step de-embedding method to accurately account for the influence of pad parasitics in silicon on-wafer RF test-structures," *IEEE Trans. Electron Devices*, vol. 48, no. 4, pp. 737–742, Aug. 2002.

[27] Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial neural networks for RF and microwave design-from theory to practice," *IEEE Trans. Microw. Theory Techn.*, vol. 51, no. 4, pp. 1339–1350, Apr. 2003.

[28] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *J. Mach. Learn. Res.*, vol. 15, no. 4, pp. 315–323, 2011.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[30] F. Wang, V. K. Devabhaktuni, C. Xi, and Q.-J. Zhang, "Neural network structures and training algorithms for RF and microwave applications," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 9, no. 3, pp. 216–240, May 1999.

[31] Y. Fang, M. C. E. Yagoub, F. Wang, and Q.-J. Zhang, "A new macromodeling approach for nonlinear microwave circuits based on recurrent neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 48, no. 12, pp. 2335–2344, 2000.

[32] 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.

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

[34] R. J. Williams and J. Peng, "An efficient gradient-based algorithm for online training of recurrent network trajectories," *Neural Comput.*, vol. 2, no. 4, pp. 490–501, Dec. 1990.

[35] G.-W. Cai, Z. Fang, and Y.-F. Chen, "Estimating the number of hidden nodes of the single-hidden-layer feedforward neural networks," in *Proc. 15th Int. Conf. Comput. Intell. Secur. (CIS)*, Dec. 2019, pp. 172–176.

Sichen Yang (Student Member, IEEE) received the B.S. degree in electronic engineering from Zhejiang University (ZJU), Hangzhou, China, in 2017, where she is currently pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering.

Her current research interest includes nonlinear modeling based on machine learning.

Xuan Chen (Student Member, IEEE) received the B.S. degree in information science and engineering from Lanzhou University, Lanzhou, China, in 2018. She is currently pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China.

Her current research interests include theoretical and applied electromagnetics and metasurface.

Chengming Wang (Student Member, IEEE) received the B.S. degree in electronic engineering from Zhejiang University, Hangzhou, China, in 2018, where he is currently pursuing the M.S. degree in electronic engineering.

His current research interests include electromagnetic compatibility (EMC) issues and nonlinear issues.

Quankun Chen (Student Member, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree.

His current research interests include nonlinear problems and electromagnetic compatibility for high-speed circuits.

Chenghan Wu (Student Member, IEEE) received the B.S. degree from Zhejiang University, Hangzhou, China, in 2019, where he is currently pursuing the master's degree.

His current research interests include device nonlinearity and contact model of a metal surface.

Zhaoyang Feng (Student Member, IEEE) is currently pursuing the Ph.D. degree with Zhejiang University, Hangzhou, China.

His current research interests include electromagnetic compatibility, signal integrity, and computational electromagnetics.

Mark D. Butala (Member, IEEE) received the H.B.E.E. degree in electrical engineering from the University of Delaware, Newark, DE, USA, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois Urbana–Champaign (UIUC), Champaign, IL, USA, in 2004 and 2010, respectively.

From 2010 to 2015, he was a Member of the Technical Staff with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. From 2016 to 2017, he was a Visiting Research Scientist with the Department of Electrical and Computer Engineering, UIUC, where he studied power transmission system impacts of geomagnetically induced currents. Since 2017, he has been with the Faculty of Zhejiang University, Hangzhou, China, where he is currently an Assistant Professor with the College of Information Science and Electronics Engineering, Zhejiang University–UIUC and an Adjunct Assistant Professor with the Department of Electrical and Computer Engineering, UIUC. His current research interests include remote sensing, image reconstruction and tomography, and statistical signal and image processing theory and applications.

Dr. Butala received a National Science Foundation Graduate Research Fellowship in 2002, a UIUC Department of Electrical and Computer Engineering Distinguished Fellowship in 2005, a UIUC Computational Science and Engineering Fellowship in 2008, and a Massachusetts Institute of Technology Lincoln Laboratories Graduate Fellowship, in 2008. He was a recipient of three NASA Group Achievement Awards.

Er-Ping Li (Fellow, IEEE) has been a Research Associate/Fellow with the School of Electronic and Information Technology, Sheffield Hallam University, Sheffield, U.K., a Senior Research Fellow, a Principal Research Engineer, an Associate Professor, and the Technical Director of the Singapore Research Institute and Industry, Singapore, since 1989. In 2000, he joined the Singapore A∗STAR Research Institute of High Performance Computing, Singapore, as a Principal Scientist and the Director of the Electronic and Photonics Department. He is currently a Changjiang-Qianren Distinguished Professor with the Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China, and the Dean of the Joint Institute of Zhejiang University–University of Illinois at Urbana–Champaign, Champaign, IL, USA. He has authored or coauthored more than 400 articles published in the referred international journals and conferences and authored two books published by John-Wiley–IEEE Press and Cambridge University Press. He holds and has filed a number of patents at the U.S. patent office. His current research interests include electrical modeling and design of microscale/nanoscale integrated circuits, 3-D electronic package integration, and nanoplasmonic technology.