Progress In Electromagnetics Research M, Vol. 120, 167–178, 2023

Analytical Neuro-Space Mapping Technology for Heterojunction Bipolar Transistors Modeling

Shuxia Yan[1, 2], Yuxing Li[1], Chenglin Li[1], Fengqi Qian[1], Xu Wang[1, 2, *], and Wenyuan Liu[3]

Abstract—An analytical modeling method for heterojunction bipolar transistor (HBT) is proposed in this paper. The new neuro-space mapping (Neuro-SM) model applied to DC, small signals and large signals simultaneously consists of two mapping networks, which provide the additional degrees of freedom. Sensitivity analysis expressions are derived to accelerate the training process. When the nonlinearity of device is high, or the response of the model is complex, the weights in the proposed model are automatically adjusted to address the accuracy limitations. The proposed modeling method is verified by measured HBT examples in DC, small signals and large signals Harmonic Balance (HB) simulation. The modeling experiments of the measured HBT demonstrate that the errors of the proposed Neuro-SM model are less than 2% by matching combined DC, small-signal S-parameters and large-signal HB data, which are less than the errors of the traditional Neuro-SM model and the coarse model. The proposed analytic Neuro-SM model fits the response of the fine model well.

1. INTRODUCTION

Nowadays, heterojunction bipolar transistors (HBTs) with high-performance are widely used in large-scale integrated circuits [1–3]. In high performance and reliability circuit/system design, accurate HBTs models play an important role [4, 5]. Efficient HBT models greatly improve circuit performance and shorten the design cycle. With the continuous improvement of monolithic circuit integration, researches on new modeling methods improving the accuracy of the existing models are needed.

Many researchers made contributions to HBT modeling methods [6–10]. Some existing traditional modeling techniques such as equivalent circuit modeling match HBT data by manually adjusting the parameters, which are heavily based on trial-and-error process [11]. For the high accuracy design requirements, these models often fail to be used directly due to the lack of freedom. The physical modeling approaches which are widely used in new devices modeling become essential to achieve design accuracy [12]. However, physical simulation requires more information about material parameters and device geometric structure. The challenge on expensive computational cost must be considered in more complex HBT modeling. Artificial neural networks (ANNs) are considered to be an effective alternative to traditional modeling technology, which learn the characteristic without the internal information of devices [13–15]. ANN was used in InP HBT small signals modeling [16].

Recently, neuro-space mapping (Neuro-SM) with the advantages of space mapping and neural network was applied to transistor modeling [17, 18]. Neuro-SM could represent the nonlinear relationship between the existing model and the modeling data by the mapping networks. The trained Neuro-SM model is embedded in circuit simulation modeling easily [19, 20]. The Neuro-SM technique in [21]

Received 7 August 2023, Accepted 25 October 2023, Scheduled 7 November 2023

* Corresponding author: Xu Wang (jiaxuwang@tju.edu.cn).

[1] School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China. [2] School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China. [3] School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China.

was used in HBT modeling. Input mapping networks are used into the model to adjust the current and voltage signals, making the Neuro-SM model match the DC data of device accurately. In order to increase the complexity and accuracy of the existing models, several improvements on Neuro-SM modeling method are subsequently studied [22]. The Neuro-SM formulations for small-signal simulation are derived to minimize the difference between the existing model and new HBT data [23]. The simulation results show that the Neuro-SM model with appropriate weights could present the HBT device. The circuit-based Neuro-SM model in [24] is developed for large-signal Harmonic Balance (HB) simulation. The perturbation sensitivity analysis in training makes modeling process time-consuming. In addition, when higher accuracy requirements of modeling are needed, or worse existing models are available, existing mapping structures are insufficient.

In this paper, an accurate analytical Neuro-SM model combining the input and output mapping neural networks is proposed. Sensitivity formulations of the new model are derived. Simultaneously, a new training method is proposed to automatically adjust the weights in neural networks. By this approach, the analytical Neuro-SM model is more accurately and efficiently applied to the DC, small-signal and large-signal simulation. The modeling experiments of the measured HBT demonstrate that the new modeling technology has high accuracy and efficiency.

2. SENSITIVITY ANALYSIS OF THE PROPOSED NEURO-SM TECHNOLOGY

In this paper, the coarse model represents the available equivalent circuit models, while the fine model represents the data from actual measurements or software simulators. The proposed Neuro-SM modeling method is to build the relationship between an accurate fine model and an imprecise coarse model by two mapping networks. Let I_b and V_c represent the HBT input signals. Let V_b and I_c represent the HBT output signals. The subscripts c and f, which represent the coarse model and the fine model, are added in the variables. The circuit structure of the HBT model is shown in Figure 1. The input signals of the fine model are tuned through mapping networks instead of acting directly on the coarse model. During the proposed modeling process, sensitivity analysis is required to provide gradient information. Gradient information can effectively improve the training efficiency and accuracy. Brute-force perturbation which reduces training speed will be performed if effective gradient information is unavailable. In order to establish an efficient modeling method, the sensitivity of the Neural-SM model is analyzed in this paper.

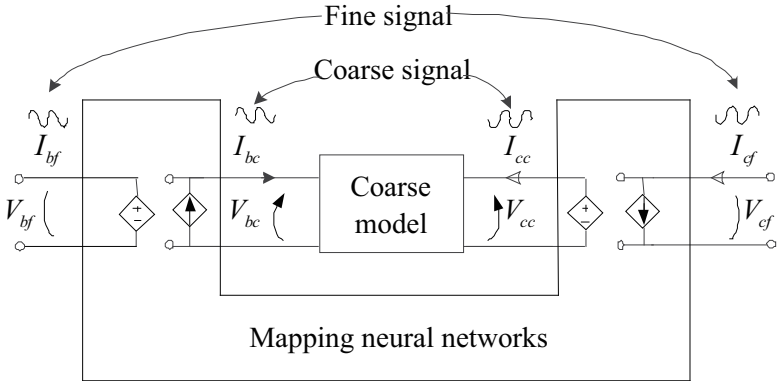


Figure 1. The circuit structure of the new Neuro-SM model.

2.1. Sensitivity Analysis of the DC Characteristic

If the unknown relationship between the coarse and fine models is highly nonlinear, it is impossible to overcome the gap between them only by changing the variables' values in the coarse model. In order to improve the existing model, two highly flexible mapping networks with sensitivity formulations are added to the coarse model. Figure 2 shows the DC signal flowchart of the proposed Neuro-SM model. The input signals $[I_{bf}, V_{cf}]^T$ are mapped to $[I_{bc}, V_{cc}]^T$ by the input network f_{ANN}. The output signals

$[V_{bc},\ I_{cc}]^T$ are mapped to $[V_{bf},\ I_{cf}]^T$ by the output network h_{ANN}. Both f_{ANN} and h_{ANN} can be implemented using multi-layer perceptrons, where $\mathbf{w}_1$ and $\mathbf{w}_2$ represent the weights in the input and output networks, respectively.

The sensitivity analysis of the model in Figure 2 is derived to find the optimal weight value of the model more quickly and efficiently. We define $w_{1,i}$ as the ith component in the weight $\mathbf{w}_1$ of the network f_{ANN} and $w_{2,j}$ as the jth component in the weight $\mathbf{w}_2$ of the network h_{ANN}. Let N_h and N_k represent the maximum number of hidden neurons in the networks f_{ANN} and h_{ANN}, respectively. The sensitivity formulations of the DC characteristic are proposed as

$$\begin{aligned}\frac{\partial\left(V_{bf}, I_{cf}\right)}{\partial w_{1,i}} &= \left(\frac{\partial\left(V_{bf}, I_{cf}\right)}{\partial\left(V_{bc,}, I_{cc}\right)}\right)^T \cdot \left(\frac{\partial\left(V_{bc}, I_{cc}\right)}{\partial\left(I_{bc}, V_{cc}\right)}\right)^T \cdot \frac{\partial\left(I_{bc}, V_{cc}\right)}{\partial w_{1,i}} \\ &= \left(\frac{\partial h_{ANN}^T\left(I_{bf}, V_{cf}, V_{bc,}, I_{cc}, \mathbf{w}_2\right)}{\partial\left(V_{bc}, I_{cc}\right)}\right)^T \cdot G_c \cdot \frac{\partial f_{ANN}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}{\partial w_{1,i}} \end{aligned} \tag{1}$$

and

$$\frac{\partial\left(V_{bf}, I_{cf}\right)}{\partial w_{2,j}} = \frac{\partial h_{ANN}\left(I_{bf,}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2\right)}{\partial w_{2,j}} \tag{2}$$

where $G_c = (\partial(V_{bc}, I_{cc})/\partial(I_{bc}, V_{cc}))^T$ is the DC transconductance matrix of the coarse model. $(\partial h_{ANN}^T(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2)/\partial(V_{bc}, I_{cc}))^T$ is the first-order partial derivative of the output with respect to the input of the network h_{ANN}. $\partial f_{ANN}(I_{bf}, V_{cf}, \mathbf{w}_1)/\partial w_{1,i}$ and $\partial h_{ANN}(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2)/\partial w_{2,j}$ are the first-order partial derivatives of f_{ANN} and h_{ANN} with respect to the weights $w_{1,i}$ and $w_{2,j}$.

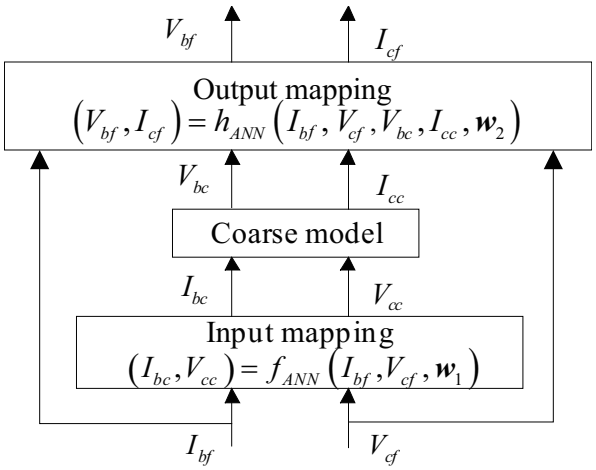


Figure 2. Flowchart of the DC signals in the proposed model.

2.2. Sensitivity Analysis of the Small Signal Characteristic

The small signal characteristics of HBT are expressed by the H-parameter. The small signal model for HBT includes the conversion between S-parameter and H-parameter. The H-matrix of the fine model H_f is calculated by the H-matrix H_c and several matrices representing the mapping network relationship. Thus, the analytical H-parameter expression of the proposed model is derived as

$$\mathbf{H}_f = \left[\begin{array}{cc} \dfrac{\partial V_{bf}}{\partial I_{bf}} & \dfrac{\partial V_{bf}}{\partial V_{cf}} \\ \dfrac{\partial I_{cf}}{\partial I_{bf}} & \dfrac{\partial I_{cf}}{\partial V_{cf}} \end{array}\right] \Bigg|_{(V_{bf}, I_{cf}) = h_{ANN}\left(I_{bf}, V_{cf}, V_{bc}, I_{cc}\big|_{(I_{bc}, V_{cc}) = f_{ANN}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}, \mathbf{w}_2\right)}$$

$$
\begin{aligned}
= & \left(\frac{\partial h_{ANN}^{T}\left(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2\right)}{\partial\left(V_{bc}, I_{cc}\right)^{T}} \Bigg|_{(V_{bc},I_{cc})=(V_{bc},I_{cc})} \Bigg|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \right)^{T} \\
& \cdot \mathbf{H}_c \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \cdot \left(\frac{\partial f_{ANN}^{T}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}{\partial\left(I_{bf}, V_{cf}\right)^{T}} \right)^{T} \\
& + \left(\frac{\partial h_{ANN}^{T}\left(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2\right)}{\partial\left(I_{bf}, V_{cf}\right)^{T}} \Bigg|_{(V_{bc},I_{cc})=(V_{bc},I_{cc})} \Bigg|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \right)^{T}
\end{aligned} \tag{3}
$$

where $\boldsymbol{H}_f$ is evaluated at the bias point (V_{bf}, I_{cf}). The derivatives of f_{ANN} and h_{ANN} at the bias point (I_{bf}, V_{cf}) are calculated by the adjoint neural network method. $\boldsymbol{H}_c$ is a matrix which contains all the internal calculation relationships of the coarse model.

The analytical small-signal expression in Equation (3) includes the influence of the mapping networks f_{ANN} and h_{ANN}, making the response of the Neuro-SM model consistent with the fine model. To ensure model consistency, the same weight values $\mathbf{w}_1$ and $\mathbf{w}_2$ are used in the mapping networks. The small signal sensitivity formulations of the mapping networks f_{ANN} and h_{ANN} with respect to weight $w_{1,i}$ and $w_{2,j}$ are proposed as

$$
\begin{cases}
\dfrac{\partial \mathbf{H}_f}{\partial w_{1,i}} = \left(\mathbf{M} \cdot H_c \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \cdot \dfrac{\partial f_{ANN}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}{\partial w_{1,i}} \right) \\
\cdot \mathbf{H}_c \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \cdot \left(\dfrac{\partial f_{ANN}^{T}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}{\partial\left(I_{bf}, V_{cf}\right)^{T}} \right)^{T} \\
+\boldsymbol{N} \cdot \left(\dfrac{\partial \mathbf{H}_c}{\partial\left(V_{bc}, I_{cc}\right)} \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \cdot \dfrac{\partial f_{ANN}\left(I_{bf}, V_{ff}, \mathbf{w}_1\right)}{\partial w_{1,i}} \right) \cdot \left(\dfrac{\partial f_{ANN}^{T}\left(I_{bf}, V_{ff}, \mathbf{w}_1\right)}{\partial\left(I_{bf}, V_{cf}\right)^{T}} \right)^{T} \\
+ \boldsymbol{N} \cdot \ \mathbf{H}_c|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \cdot \left(\dfrac{\partial^2 f_{ANN}^{T}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}{\partial\left(I_{bf}, V_f\right)^{T} \partial w_1} \right)^{T} \\
\mathbf{M} = \left(\dfrac{\partial^2 h_{ANN}^{T}\left(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2\right)}{\partial\left(V_{bc}, I_{cc}\right)^{T} \partial\left(V_{bc}, I_{cc}\right)} \Big|_{(V_{bc},I_{cc})=(V_{bc},I_{cc})} \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \right)^{T} \\
\mathbf{N} = \left(\dfrac{\partial h_{ANN}^{T}\left(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2\right)}{\partial\left(V_{bc}, I_{cc}\right)^{T}} \Big|_{(V_{bc},I_{cc})=(V_{bc},I_{cc})} \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \right)^{T}
\end{cases} \tag{4}
$$

and

$$
\begin{aligned}
\frac{\partial \mathbf{H}_f}{\partial w_{2,j}} = & \left(\frac{\partial^2 h_{ANN}^{T}\left(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2\right)}{\partial\left(V_{bc}, I_{cc}\right)^{T} \partial w_{2,j}} \Bigg|_{(V_{bc},I_{cc})=(V_{bc},I_{cc})} \Bigg|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \right)^{T} \\
& \cdot \mathbf{H}_c \Big|_{(I_{bc},V_{cc})=f_{ANN}\left(I_{bf},V_{cf},\mathbf{w}_1\right)} \cdot \left(\frac{\partial f_{ANN}^{T}\left(I_{bf}, V_{cf}, \mathbf{w}_1\right)}{\partial\left(I_{bf}, V_{cf}\right)^{T}} \right)^{T}
\end{aligned} \tag{5}
$$

respectively. In Equation (4), $(\partial^2 h_{ANN}^{T}(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2)/\partial(V_{bc}, I_{cc})^{T}\partial(V_{bc}, I_{cc}))^{T}$ represents the second-order partial derivative of the mapping network h_{ANN}, which is the derivative of the Jacobian matrix $(\partial h_{ANN}^{T}(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2)/\partial(V_{bc}, I_{cc})^{T})^{T}$ with respect to (V_{bc}, I_{cc}). $(\partial^2 f_{ANN}^{T}(I_{bf}, V_{cf}, \mathbf{w}_1)/\partial(I_{bf}, V_{cf})^{T}\partial w_{1,i})^{T}$ represents the second-order partial derivative of the network

f_{ANN}, which is the derivative of the matrix $(\partial f^T_{ANN}(I_{bf}, V_{cf}, \mathbf{w}_1)/\partial(I_{bf}, V_{cf})^T)^T$ with respect to weight $w_{1,i}$. $(\partial f^T_{ANN}(I_{bf}, V_{cf}, \mathbf{w}_1)/\partial(I_{bf}, V_{cf})^T)^T$ represents the first-order partial derivative of the network f_{ANN}, which is the derivative of $\partial f^T_{ANN}(I_{bf}, V_{cf}, \mathbf{w}_1)$ with respect to $\partial(I_{bf}, V_{cf})^T$. In Equation (11), $(\partial^2 h^T_{ANN}(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2)/\partial(V_{bc}, I_{cc})^T \partial w_{2,j})^T$ represents the second-order partial derivative of the network h_{ANN}, which is the derivative of the matrix $(\partial h^T_{ANN}(I_{bf}, V_{cf}, V_{bc}, I_{cc}, \mathbf{w}_2)/\partial(V_{bc}, I_{cc})^T)^T$ parameters with respect to weight $w_{2,j}$.

2.3. Sensitivity Analysis of the Large Signal Characteristic

The large-signal output of the Neuro-SM model is derived within the environment of HB simulation. The time domain signals are used in the mapping networks and the coarse model, while the frequency domain signals are needed in the fine model. Inverse fast Fourier transform and fast Fourier transform are introduced into the large signal model. In the fine model, the harmonic current and voltage signals are defined as $I_{cf}(\omega_k)$ and $V_{bf}(\omega_k)$. In the coarse model, the harmonic current and voltage signals are defined as $I_{cc}(\omega_k)$ and $V_{bc}(\omega_k)$. The subscript k represents the harmonic frequency index. The value of k ranges from 0 to the maximum harmonics number N_H. The large-signal flowchart of the new model is shown in Figure 3.

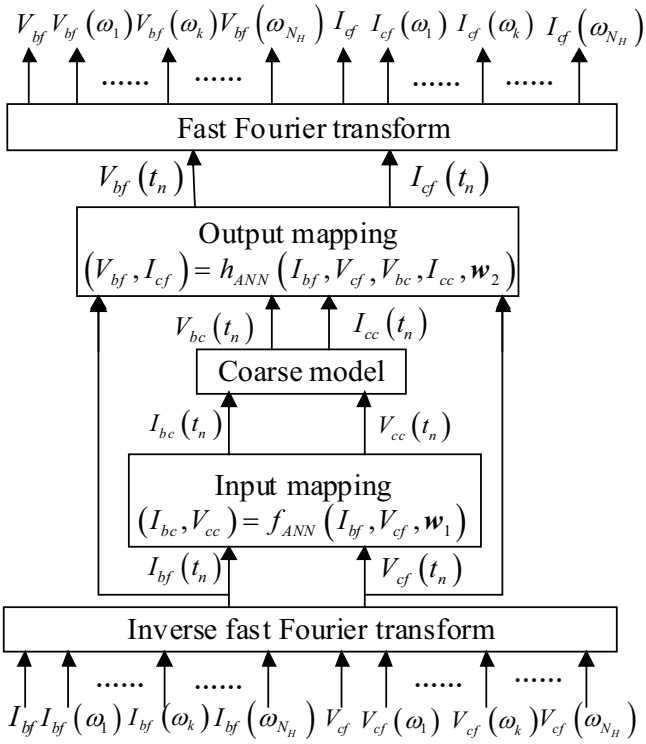


Figure 3. Flowchart of the large signals in the proposed model.

In the proposed large-signal model, $F()$ is used to represent the fast Fourier transform calculation. The Fourier coefficient of the kth harmonic in the nth sampling could be represented by $W_{N_T}(n,k) = e^{-j2\pi nk/N_T}$, where the subscript n is the sampling time point, and N_T is the maximum time point. The

sensitivity formulations of the mapping networks f_{ANN} and h_{ANN} are proposed as

$$\begin{aligned}&\frac{\partial\left(V_{bf}\left(\omega_k\right),I_{cf}\left(\omega_k\right)\right)}{\partial w_{1,i}}\\ &=\frac{1}{N_T}\sum_{n=0}^{N_T-1}\frac{\partial h_{ANN}\left(I_{bf}\left(t_n\right),V_{cf}\left(t_n\right),V_{bc}\left(t_n\right),I_{cc}\left(t_n\right)\Big|_{\left(I_{bc}(t_n),V_{cc}(t_n)\right)=f_{ANN}\left(I_{bf}(t_n),V_{cf}(t_n),\mathbf{w}_1\right)},\mathbf{w}_2\right)}{\partial\left(V_{bc}\left(t_n\right),I_{cc}\left(t_n\right)\right)}\\ &\cdot G_c\cdot\frac{\partial f_{ANN}\left(I_{bf}\left(t_n\right),V_{cf}\left(t_n\right),\mathbf{w}_1\right)}{\partial w_{1,i}}\cdot W_{N_T}\left(n,k\right)\end{aligned}\tag{6}$$

and

$$\begin{aligned}&\frac{\partial\left(V_{bf}\left(\omega_k\right),I_{cf}\left(\omega_k\right)\right)}{\partial w_{2,j}}\\ &=\frac{1}{N_T}\sum_{n=0}^{N_T-1}\frac{\partial h_{ANN}\left(I_{bf}(t_n),V_{cf}(t_n),V_{bc}(t_n),I_{cc}(t_n)\Big|_{\left(I_{bc}(t_n),V_{cc}(t_n)\right)=f_{ANN}\left(I_{bf}(t_n),V_{cf}(t_n),\mathbf{w}_1\right)},\mathbf{w}_2\right)}{\partial w_{2,j}}\cdot W_{N_T}(n,k)\end{aligned}\tag{7}$$

where $G_c=(\partial(V_{bc}(t_n),I_{cc}(t_n))/\partial(I_{bc}(t_n),V_{cc}(t_n)))^T$ represents the large-signal transconductance matrix at time t_n, which is calculated at the bias point $(I_{bc}(t_n),V_{cc}(t_n))$.

2.4. The Proposed Training Method for the Analytical Neuro-SM Model

In this paper, the proposed training method utilizes the input and output neural networks to narrow the gap between the characteristics of the modeling device and the coarse model, making the proposed model represent the device data accurately. The training is mainly done with the data from the simulators or measured by device equipment. Training error is represented as

$$\begin{aligned}E\left(\mathbf{w}_1,\mathbf{w}_2\right)\;=\;&\frac{1}{2}\sum_{q=1}^{N_{V_{cf}}}\sum_{p=1}^{N_{I_{bf}}}\left\|A\left(D\left(I_{Bp},V_{Cq},\mathbf{w}_1,\mathbf{w}_2\right)-D_{Fp}^{q}\right)\right\|^2\\ &+\frac{1}{2}\sum_{q=1}^{N_{V_{cf}}}\sum_{p=1}^{N_{I_{bf}}}\sum_{l=1}^{N_{freq}}\left\|B\left(S\left(I_{Bp},V_{Cq},\omega_k,\mathbf{w}_1,\mathbf{w}_2\right)-S_{Fpq}^{l}\right)\right\|^2\\ &+\frac{1}{2}\sum_{k=1}^{N_H}\sum_{q=1}^{N_{V_{cf}}}\sum_{p=1}^{N_{I_{bf}}}\left\|C\left(H\left(I_{Bp},V_{Cq},\omega_k,\mathbf{w}_1,\mathbf{w}_2\right)-H_{Fpq}\left(\omega_k\right)\right)\right\|^2\end{aligned}\tag{8}$$

where $D(\cdot)$, $S(\cdot)$ and $H(\cdot)$ indicate the DC, small-signal and the large-signal output of the Neuro-SM model, respectively. D_F, S_F and $H_F(\cdot)$ indicate the DC, the small-signal and the large-signal output of the fine model. In order to find the optimal value easily, we add three scaling variables into the error formula, i.e., A, B, and C. The subscripts $p(p=1,2,\ldots,N_{I_{bf}})$ and $q(q=1,2,\ldots,N_{V_{cf}})$ represent the pth and qth input of the training data, respectively. $N_{I_{bf}}$ and $N_{V_{cf}}$ are the maximum numbers of the bias data. The subscript $l(l=0,1,2,\ldots,N_{freq})$ and $k(k=0,1,2,\ldots,N_H)$ represent the lth input frequency and the kth harmonic frequency. N_{freq} is the maximum frequency of the small-signal data, and N_H is the maximum harmonic of the large-signal data.

Figure 4 shows the proposed training flowchart of the model with two mapping networks. The mapping networks f_{ANN} and h_{ANN} are trained to minimize the training error as shown in Equation (8). The errors are expressed as the gap between the characteristics of the modeling device and the coarse model. The training process is summarized as follows: Unit mapping networks, in which the output is equal to the input, are constructed to prevent the model accuracy from dropping. Then, the weights $\mathbf{w}_1$ and $\mathbf{w}_2$ in f_{ANN} and h_{ANN} are optimized with the sensitivity analysis equations, respectively. The criterion is that the training error is less than the threshold. If the training error in Equation (8) meets

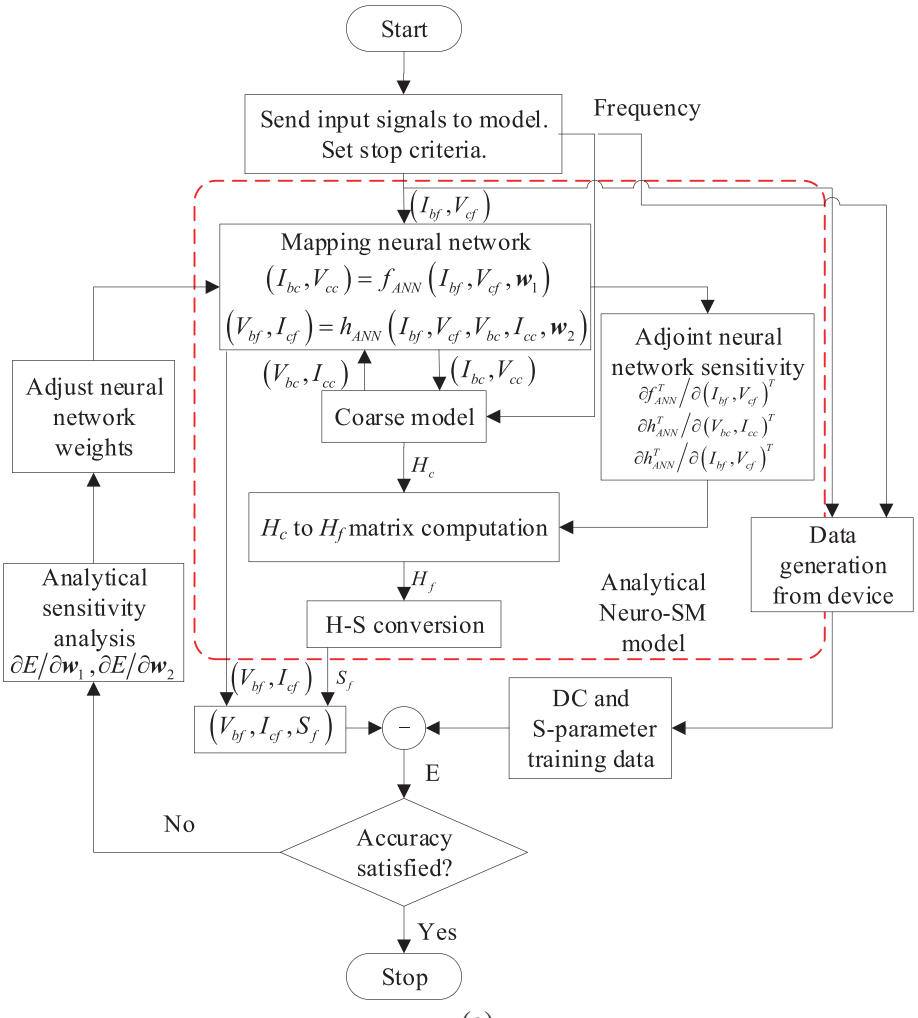
the accuracy requirements, the optimization process is over. After training, the proposed Neuro-SM model can be tested by the test data, which include the DC, small-signal S-parameter, and large-signal HB data.

3. EXAMPLES

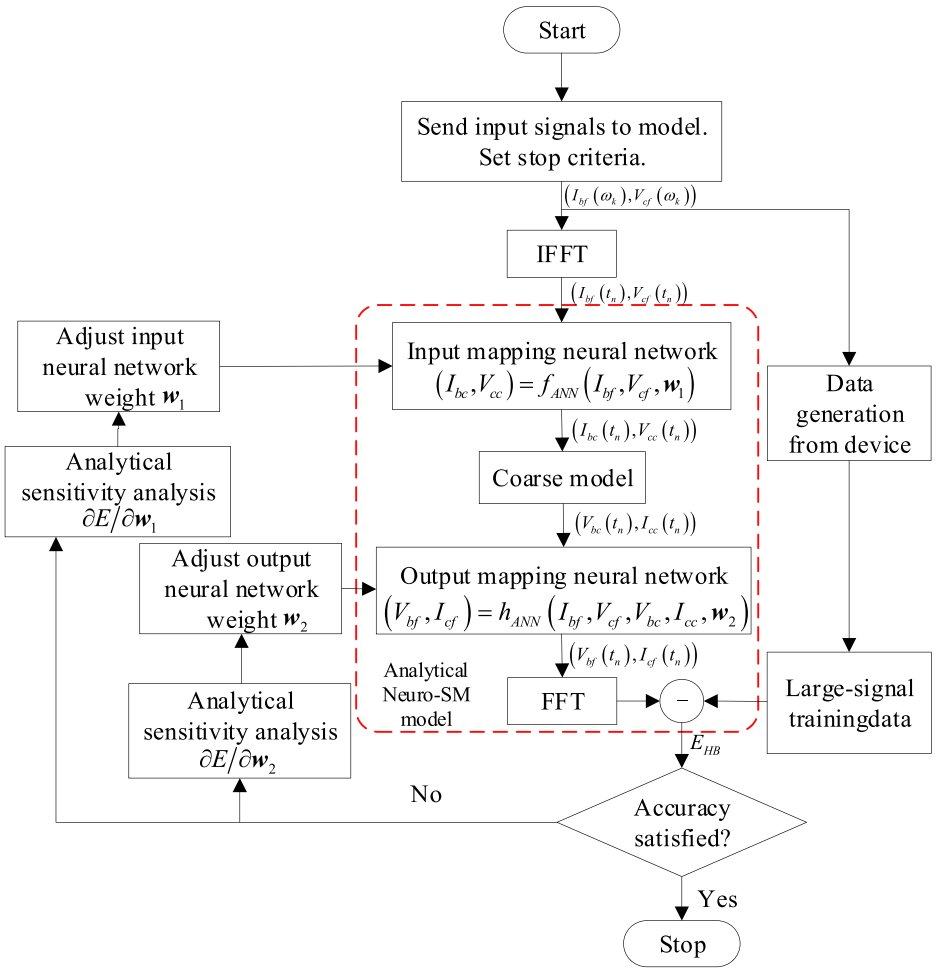
In this section, the modeling method with the sensitivity analysis is used to develop an accurate model for the InGaP HBT [23]. The on-chip measurements data is used as the fine model, while the coarse model is the Agilent HBT model. The Agilent HBT model could not match the measurements data well, even after improving the characteristic as much as possible. We implement the new model in NeuroModelerPlus [25]. The proposed Neuro-SM model with sensitivity analysis is trained and tested with the on-chip measurements data.

The input unit mapping training was done at 4680 different bias points in the following range: I_{bf} is from 10 µA to 200 µA while the step is 5 µA, and V_{cf} is from 0.05 V to 6 V while the step is 0.05 V. The output unit mapping training was done at 5280 different bias points in the following range: I_{bf} is from 5 µA to 220 µA while the step is 5 µA, and V_{cf} is from 0.05 V to 6 V while the step is 0.05 V. The hidden neurons in the networks f_{ANN} and h_{ANN} are 40 and 50.

Table 1 shows the ranges of the training data and test data for modeling in the formal training. The proposed Neuro-SM model is trained using DC data at 250 different bias points and small-signal data at 9 different bias points for 423 training iterations. When large-signal harmonic simulation is conducted, the highest harmonic order is set as 5, while the first three harmonics are taken as the training data.



(a)



(b)

Figure 4. The training flowchart of the proposed model. (a) DC and small-signal training process. (b) Large-signal training process.

Table 1. Ranges of the training data and test data for modeling.

Simulation type	Input variable	Training data	Test data
DC simulation	I_{bf} (mA)	0.02 : 0.04 : 0.18	0.04 : 0.04 : 0.16
	V_{cf} (V)	0.05 : 0.1 : 4.95	0.1 : 0.1 : 4.9
S-parameter simulation	I_{bf} (mA)	0.05 : 0.02 : 0.09	0.06 : 0.02 : 0.08
	V_{cf} (V)	1.4 : 0.25 : 1.9	1.45 : 0.4 : 1.85
	freq (GHz)	0.05 : 0.05 : 0.5; 0.6 : 0.1 : 1; 1.2 : 0.2 : 3; 3.25 : 0.25 : 4.5; 5 : 1 : 20	
HB simulation	I_{bf} (mA)	0.06 : 0.03 : 0.09	0.07 : 0.01 : 0.08
	V_{cf} (V)	1.4 : 0.4 : 1.8	1.5
	RF power (dBm)	−19 : 1 : 15	−19 : 1 : 15

To verify the feasibility and accuracy of the proposed analytical Neuro-SM model, the traditional Neuro-SM method mentioned in [21] is used for this example. In contrast to the traditional method that incorporates neural networks only at the input, the proposed model adds networks at both the input and output, which increases the degree of freedom and improves the accuracy of the model. Table 2 shows the error of the proposed Neuro-SM model, traditional Neuro-SM model, and coarse model by matching combined DC, small-signal S-parameter, and large-signal harmonic data.

Table 2. Modeling results of the three models.

Model type	DC matching	S-parameter matching	Large-signal matching	Combined DC, S-parameter, large-signal matching
Coarse model	2.1437%	9.5896%	9.7447%	22.7246%
Traditional model	0.7443%	4.5174%	2.3416%	7.418%
Proposed model	0.4772%	1.4236%	1.6947%	1.9261%

Figure 5 shows the output voltage V_{bf} and output current I_{cf} of the coarse model, traditional Neuro-SM model, proposed Neuro-SM model, and fine model in the DC simulation. There is a certain gap between the coarse model and fine model. Even though the traditional Neuro-SM model and proposed Neuro-SM model match the fine model well, the proposed analytical Neuro-SM model has higher modeling accuracy than the existing models in detail.

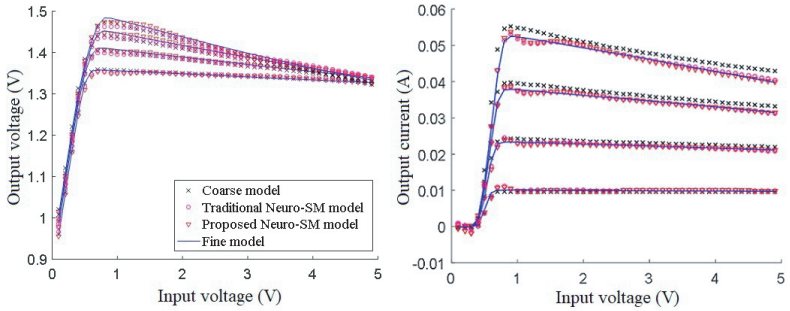


Figure 5. The DC results of the four models. (a) Output voltage characteristics. (b) Output current characteristics.

The S-parameter results of the fine model, proposed Neuro-SM model, traditional Neuro-SM model, and coarse model are shown in Figure 6. The proposed analytical Neuro-SM model fits the fine model well, while the traditional Neuro-SM model has a large gap with the fine model, especially in the real and imaginary parts of the S_{12}. The proposed analytical Neuro-SM model can match the fine model in both the curve trend and the details. It verifies that the new model has high accuracy and wide application in small-signal S-parameter simulation.

Figure 7 shows the large-signal harmonic results of the coarse model, traditional Neuro-SM model, proposed Neuro-SM model, and fine model. It can be seen from Figure 7 that the proposed analytical Neuro-SM model fits the fine model well at the first three harmonics. As the power increases, the response of the traditional Neuro-SM model cannot be consistent with that of the fine model. The proposed analytic Neuro-SM model fits the response of the fine model within the training range.

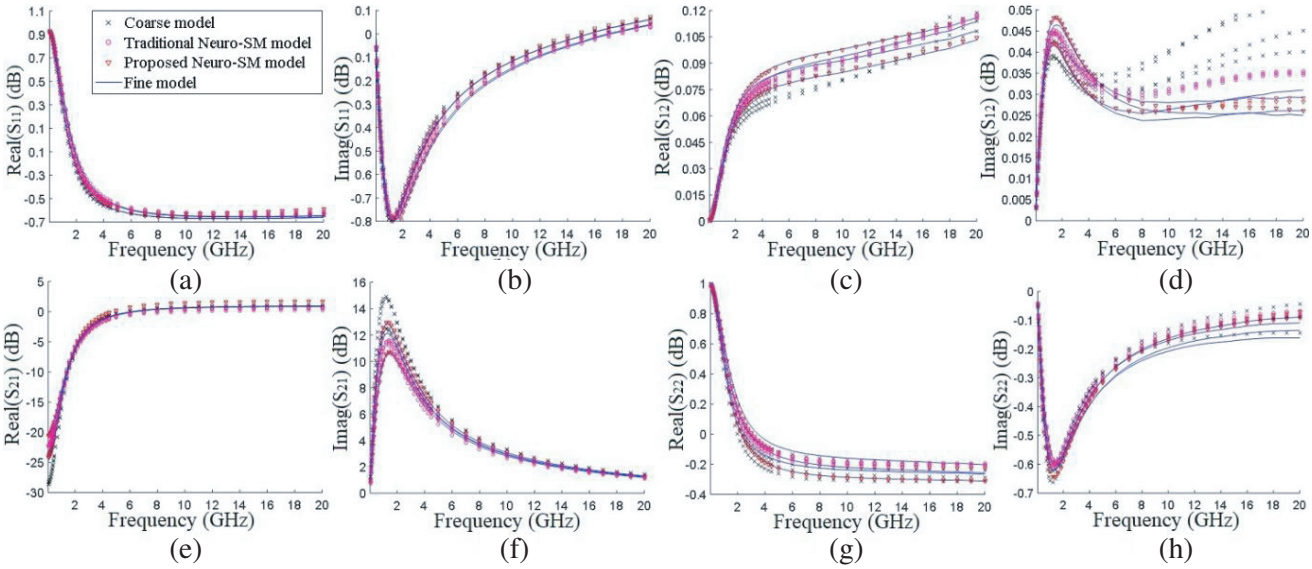


Figure 6. The S-parameter results of the four models. (a) Real(S_{11}). (b) Imag(S_{11}). (c) Real(S_{12}). (d) Imag(S_{12}). (e) Real(S_{21}). (f) Imag(S_{21}). (g) Real(S_{22}). (h) Imag(S_{22}).

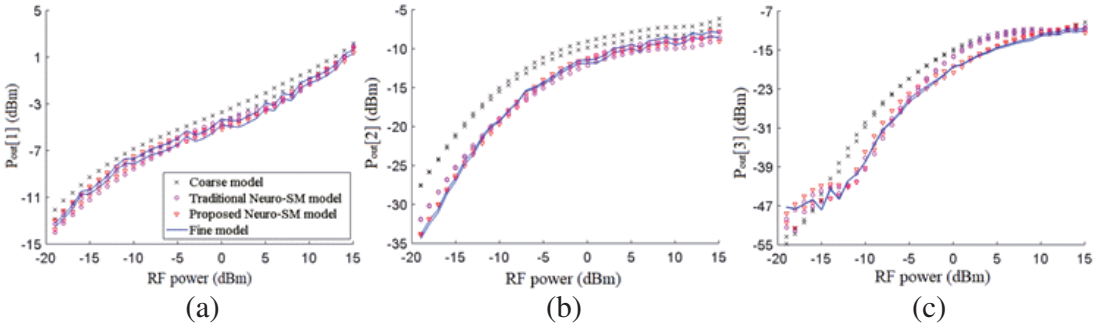


Figure 7. The large-signal results of the four models. (a) Pout[1]. (b) Pout[2]. (c) Pout[3].

We can see from Figure 5, Figure 6, and Figure 7 that the accuracy of the traditional Neuro-SM model is better than that of the coarse model. The input mapping network in the traditional Neuro-SM model could change the weights to minimize the performance gap between the fine model and coarse model. However, when the non-linearity of device is high, the traditional model can no longer meet the accuracy requirements. The proposed analytical Neuro-SM model combining the coarse model and the two networks greatly improves the model accuracy and matches the fine data well. The mapping network h_{ANN} adding on the coarse model provides necessary degrees of freedom to improve the proposed model accuracy.

In addition, the simpler sensitivity formulations greatly improve the efficiency of the training process. The circuit-based Neuro-SM method with perturbation analysis in [22] is used as a comparison. Table 3 shows the training CPU time, which confirms that the proposed training algorithm has better efficiency. Circuit-based Neuro-SM is implemented by controlled sources, which introduces additional variables and equations into circuit simulation to complicate the solution process. The proposed method has high simulation efficiency without extra variables. In addition, sensitivity analysis provides gradient information to speed up training process. The more data that is trained, the more significant the efficiency advantages of the proposed model become.

Table 3. Training time comparison of the two models.

Data	Circuit-based model	Proposed model
15 sets	20.1 s	3.2 s
35 sets	81.5 s	10.4 s
55 sets	112.9 s	14.6 s

4. CONCLUSIONS

In this paper, the analytical HBT modeling method applied to DC, small signals and large signals simultaneously is proposed for the first time. Two mapping networks are used to minimize the gap between the existing coarse model and fine HBT model. The analytical expressions combining the coarse model and mapping networks are proposed to improve the accuracy of the Nuero-SM model. The simpler sensitivity formulations greatly improve the efficiency of the training process. Compared with the circuit-based Neuro-SM method, the proposed training algorithm can save about 87% of the time cost with 55 sets of data. The measured HBT examples verify that the proposed modeling method has better accuracy and efficiency than the existing modeling methods.

ACKNOWLEDGMENT

This work was supported by the China Postdoctoral Science Foundation under Grant No. 2020M680883 and the National Natural Science Foundation of China under Grant No. 62201335.

REFERENCES

1. Squartecchia, M., T. K. Johansen, J. Y. Dupuy, et al., "E-band indium phosphide double heterojunction bipolar transistor monolithic microwave-integrated circuit power amplifier based on stacked transistors," *Microwave and Optical Technology Letters*, Vol. 61, No. 2, 550–555, 2019.
2. Boulgheb, A., M. Lakhdara, and S. Latreche, "Improvement of the self-heating performance of an advanced SiGe HBT transistor through the Peltier effect," *IEEE Transactions on Electron Devices*, Vol. 68, No. 2, 479–484, 2021.
3. Tanaka, S., "A study on AM-AM/PM characteristics of a single-stage HBT power amplifier," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E104-A, No. 2, 484–491, 2021.
4. Mohammadi, F. and A. Sadrossadat, "Modeling and simulation techniques for microwave components," *Microwave Systems and Applications*, 2017, DOI: 10.5772/66356.
5. Zhang, H., G. Niu, M. B. Willemsen, and A. J. Scholten, "Improved compact modeling of SiGe HBT linearity with MEXTRAM," *IEEE Transactions on Electron Devices*, Vol. 68, No. 6, 2597–2603, 2021.
6. Karimi, G., R. Banitalebi, and S. B. Sedaghat, "Simulation of SiGe:C HBTs using neural network and adaptive neuro-fuzzy inference system for RF applications," *International Journal of Electronics*, Vol. 100, No. 7, 959–975, 2013.
7. Rudolph, M., "Compact HBT modeling: status and challenges," *IEEE MTT-S International Microwave Symposium*, 1206–1209, Anaheim, CA, USA, 2010.
8. Johansen T. K., M. Rudolph, T. Jensen, et al., "Small- and large-signal modeling of InP HBTs in transferred-substrate technology," *International Journal of Microwave and Wireless Technologies*, Vol. 6, Nos. 3–4, 243–251, 2014.
9. Zhang, A. and J. Gao, "An improved small signal model of InP HBT for millimeter-wave applications," *Microwave and Optical Technology Letters*, Vol. 63, No. 8, 2160–2164, 2021.

10. Zhang, J., M. Liu, J. Wang, and K. Xu, "An analytic method for parameter extraction of InP HBTs small-signal model," *Circuit World*, Vol. 48, No. 4, 393–400, 2021.
11. Cheng, L., H. Lu, M. Xia, et al., "An augmented small-signal model of InP HBT with its analytical-based parameter extraction technique," *Microelectronics Journal*, Vol. 121, 105366, 2022.
12. Zhang, Q. J. and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Artech House, Norwood, Mass, 2000.
13. Feng, F., W. Na, J. Jin, et al., "Artificial neural networks for microwave computer-aided design: The state of the art," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 11, No. 70, 4597–4619, 2022.
14. Zlatica, D. M., G. Crupi, A. Caddemi, et al., "A review on the artificial neural network applications for small-signal modeling of microwave FETs," *International Journal of Numerical Modelling Electronic Networks Devices and Fields*, Vol. 33, No. 3, e2668, 2020.
15. Feng, F., W. Na, J. Jin, et al., "ANNs for fast parameterized EM modeling: the state of the art in machine learning for design automation of passive microwave structures," *IEEE Microwave Magazine*, Vol. 22, No. 10, 37–50, 2021.
16. Zhang, A. and J. Gao, "InP HBT small signal modeling based on artificial neural network for millimeter-wave application," *2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization*, 1–3, Hangzhou, China, 2020.
17. Zhu, L., J. Zhao, Z. Li, et al., "A general neuro-space mapping technique for microwave device modeling," *EURASIP Journal on Wireless Communications and Networking*, Vol. 2018, No. 1, 37, 2018.
18. Zhang, W., F. Feng, V.-M.-R. Gongal-Reddy, et al., "Space Mapping approach to electromagnetic centric multiphysics parametric modeling of microwave components," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 66, No. 7, 3169–3185, 2018.
19. Yan, S., Y. Zhang, W. Liu, et al., "A novel electromagnetic centric multiphysics parametric modeling approach using neuro-space mapping for microwave passive components," *Photonics*, Vol. 9, No. 12, 960, 2022.
20. Zhao, Z., L. Zhang, F. Feng, et al., "Space mapping technique using decomposed mappings for GaN HEMT modeling," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 68, No. 8, 3318–3341, 2020.
21. Zhang, L., J. J. Xu, M. C. E. Yagoub, et al., "Efficient analytical formulation and sensitivity analysis of neuro-space mapping for nonlinear microwave device modeling," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 53, No. 9, 2752–2767, 2005.
22. Yan, S., S. Zhang, Y. Zhang, et al., "An accurate neuro-space mapping method for heterojunction bipolar transistor modeling," *2020 IEEE MTT-S International Conference on Numerical Electromagnetic and Multiphysics Modeling and Optimization*, 1–4, Hangzhou, China, 2020.
23. Yan, S., Q. Cheng, H. Wu, and Q. J. Zhang, "Neuro-space mapping for modeling heterojunction bipolar transistor," *Transactions of Tianjin University*, Vol. 21, No. 1, 90–94, 2015.
24. Wu, H. F., Q. F. Cheng, S. X. Yan, et al., "Transistor model building for a microwave power heterojunction bipolar transistor," *IEEE Microwave Magazine*, Vol. 16, No. 2, 85–92, 2015.
25. Zhang, Q. J., "Neuro modeler plus," Dept. Electron., Carleton Univ., Ottawa, ON., Canada, 2008.