


Article

An Aging Small-Signal Model for Degradation Prediction of Microwave Heterojunction Bipolar Transistor S-Parameters Based on Prior Knowledge Neural Network

Lin Cheng¹, Hongliang Lu¹ , Silu Yan¹, Chen Liu^{1,*}, Jiantao Qiao¹, Junjun Qi¹, Wei Cheng², Yimen Zhang¹ and Yuming Zhang¹

- ¹ Key Laboratory for Wide Band Gap Semiconductor Materials and Devices of Education Ministry, School of Microelectronics, Xidian University, Xi'an 710071, China; domini_c@foxmail.com (L.C.); hllv@mail.xidian.edu.cn (H.L.); siluyan_xidian@163.com (S.Y.); 22111213679@stu.xidian.edu.cn (J.Q.); jjqi1108@163.com (J.Q.); ymzhang@xidian.edu.cn (Y.Z.); zhangym@xidian.edu.cn (Y.Z.)
- ² Science and Technology on Monolithic Integrated Circuits and Modules Laboratory, Nanjing Electronic Devices Institute, Nanjing 210016, China
- * Correspondence: liuchen@xidian.edu.cn

Abstract: In this paper, an aging small-signal model for degradation prediction of microwave heterojunction bipolar transistor (HBT) S-parameters based on prior knowledge neural networks (PKNNs) is explored. A dual-extreme learning machine (D-ELM) structure with an adaptive genetic algorithm (AGA) optimization process is used to simulate the fresh S-parameters of InP HBT devices and the degradation of S-parameters after accelerated aging, respectively. In addition to the reliability parametric inputs of the original aging problem, the S-parameter degradation trend obtained from the aging small-signal equivalent circuit is used as additional information to inject into the D-ELM structure. Good agreement was achieved between measured and predicted results of the degradation of S-parameters within a frequency range of 0.1 to 40 GHz.

Keywords: degradation prediction; S-parameters; D-ELM structure; aging small-signal equivalent circuit; prior knowledge neural network



Citation: Cheng, L.; Lu, H.; Yan, S.; Liu, C.; Qiao, J.; Qi, J.; Cheng, W.; Zhang, Y.; Zhang, Y. An Aging Small-Signal Model for Degradation Prediction of Microwave Heterojunction Bipolar Transistor S-Parameters Based on Prior Knowledge Neural Network. *Micromachines* **2023**, *14*, 2023. <https://doi.org/10.3390/mi14112023>

Academic Editors: Kuan-Wei Lee and Yeong-Her Wang

Received: 1 October 2023
Revised: 25 October 2023
Accepted: 28 October 2023
Published: 30 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid development of microelectronics technology, the increasing maturity of the integrated circuit process, and the continuous exploration of the depths of satellite communications, radar aerospace, and other fields, the application and development of RF semiconductor devices at microwave millimeter-wave frequencies have gradually occupied an important position in high-speed communications [1,2]. In order to reduce the cost and improve the performance, some key dimensions of the process have been continuously reduced. Accordingly, the electric field strength and current density to which the device is subjected are also increasing, which exacerbates the probability and randomness of the occurrence of various failure mechanisms and poses a serious challenge to the reliability of integrated circuits and the improvement of electronic design automation (EDA) [3].

High-reliability EDA tools need to fit, cope with, and correct real-world physical and process problems in software with a systematic approach and predictive margins as high as possible and ultimately ensure that the circuit design simulation results are consistent with the degradation results of the flow test in the reliability environment to achieve the optimal end-of-life performance/power consumption/area (EOL-PPA). The device model, as an important bridge between process and circuit design, is a key link in the DTCO flow under the new architecture, and its accuracy directly determines the accuracy and efficiency of circuit design [4]. Through the establishment of reliability models for key components in integrated systems, on the one hand, the internal physical mechanism

of the device can be explored in depth, laying a foundation for the realization of the device reliability degradation simulation. On the other hand, a better prediction of the characteristic degradation can be obtained in the large-scale simulation of the electrical characteristics of circuits as well as reliability degradation effects. In addition to providing guidance for process improvement, device-level reliability models also play a key role in the reliability-aware design of integrated circuits. The establishment of efficient and accurate reliability models can be used to guide the reliability design and optimization of circuits in order to reduce the time and cost of high-reliability circuit design [5].

In recent years, research on the conventional small-signal model has focused on the characterization of non-ideal effects corresponding to ultra-high frequency rates and severe parasitics [6–8]. The reliability modeling-related research on this basis has been insufficient, still focusing mainly on the physical level of the device, with constraints such as long R&D cycle time, slow modeling, and incompatibility with EDA softwares [9,10]. A reliability compact model based on semi-empirical and semi-physical approaches may lack accuracy when actually characterizing degradation in time and frequency domains due to some approximations or simplifications [11,12]. Therefore, in addition to the ongoing in-depth study of the physical nature of RF microwave transistor aging and the associated reliability of compact models, we believe that there is another approach worth exploring. This method involves the use of an artificial neural network (ANN) methodology based on machine learning-assisted behavioral-level modeling [13] to achieve degradation prediction of RF devices in harsh reliability environments. The method can be trained on the degradation behavior of the device under specific reliability conditions, and then the trained network can be used to make predictions. Although the data-driven reliability-based artificial neural network model can achieve higher accuracy, it strongly relies on a large number of reliability test samples for training and learning [14]. Even if the TCAD numerical simulation tool is used to obtain the training data, a large amount of reliable experimental data is still needed to perform the complex calibration of the degradation of the amount of marching associated with aging as a time-correlated process [15]. While microwave devices involve a variety of complex combinations of conditions such as materials, geometries, frequencies, processes, etc., reliability data are more difficult to obtain and costly, and reliability experiments require strict requirements on environment, equipment, samples, etc., which require a large amount of cost investment. It can be seen that improving the traditional reliability artificial neural network model with limited reliability test samples in order to achieve efficient, accurate, and high generalization ability modeling is particularly important. Therefore, we hope to use physical/semi-physical degradation laws or formulas to guide this behavioral model and adopt a simple and effective knowledge injection to improve the reliability of the black box model.

In this paper, a prior knowledge-based neural network approach for device reliability modeling is proposed. A dual-extreme learning machine (D-ELM) is used to decompose the aging and fresh features of the device into two sub-networks to simplify the complex nonlinear mapping of the original problem. The established degradation function of the key parameters of the device is introduced into the aging equivalent-circuit model to obtain the degradation of each component of the S -parameters. Then, the S -parameter degradation information obtained from the coarse model is injected into the degradation sub-network as prior knowledge. In addition to the reliability parameter inputs of the original aging problem, the coarse model outputs are also used as additional inputs. At this point, the input–output mapping to be learned by the neural network is the mapping between the existing approximate reliability model outputs and the original problem, thus weakening the complex mapping relationship between reliability parameters and S -parameter degradation and improving the quality of the mapping. This, in turn, reduces the dependence on reliable training samples and improves accuracy with the high generalization ability of the model. Moreover, we have utilized an adaptive evolutionary algorithm to optimize the model and improve the stability of network training. Finally, the proposed PKNN-based reliability model is validated by applying it to 0.7 μm InGaAs/InP HBTs.

2. Device Structure and Experiment

The InP DHBT device used in this paper adopts a three mesa-structure, and its structural schematic is shown in Figure 1, in which the non-alloyed Ti/Pt/Au and Pt/Ti/Pt/Au are used for N-type and P-type ohmic contacts, respectively. Base self-alignment technology, non-contact exposure, and wet etching technology are used in the fabrication of the device.

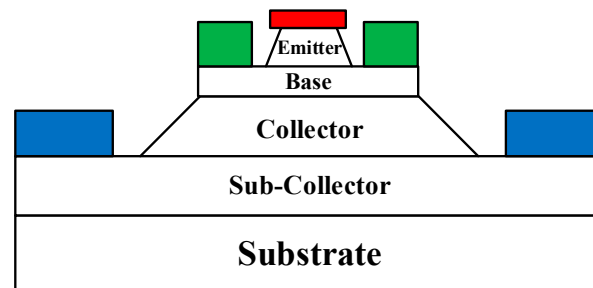


Figure 1. Device structure schematic.

In this paper, we use a Keysight B1500A semiconductor device parameter analyzer to apply high-field electrical stress to the device by means of a DC probe. The B1500A has a very sophisticated cycling control module for DC probes to precisely apply the electrical stress, and this test module can also monitor whether the junction of the InP HBT device is broken during the test. The schematic diagram of high-field electrical stress application is shown in Figure 2, in which the base and collector of the device are connected to the SMU1 and SMU2 ports of Keysight B1500A, respectively. We set the two ports to voltage input mode and used the EasyEXPERT group+ 2015 software for the control of the electrical stress. The selection of stresses needs to be carried out before conducting the high-field stress aging test. Applying a reverse high-field stress of about 80% of the device breakdown voltage BV_{CBO} to the BC junction can allow the device to operate properly without breakdown and cause non-negligible degradation [16]. We have chosen about 70–90% of the breakdown voltage for high-field stress aging. The breakdown voltage of the InP HBT device used is about 4.7 V; thus, we apply constant reverse bias voltages of 3.4–4.3 V to the BC junction of the device and select four stress biases in steps of 0.3 V to perform S-parameter tests after up to 300 minutes of stress application while keeping the BE junction open. To accurately assess the effect of high-field electrical stress on device RF characteristics, a Rohde & Schwarz ZVA 50 Network Analyzer controlled by IC-CAP 2018 software is employed for S-parameter testing, and the ground–signal–ground (GSG) microwave probe was used in the on-chip measuring process. In addition, to correct the error terms introduced by the test equipment itself, a system error calibration is required before testing with VNA. In this paper, the on-chip test of InP HBT S-parameters is a two-port network test system, and to obtain the actual S-parameters of the DUT, the TOSM (Through, Open, Short, Match) system error calibration method [17,18] is used to remove the parasitic elements introduced by the equipment components such as test cables and probe tips before the reliability test. It is worth noting that the calibration of the RF on-chip test platform is performed before the electrical stress is applied. Also, attention is paid to maintaining a stable test environment during the test to ensure that the calibration conditions do not deviate. All of the above on-chip tests are performed on the CASCADE MICROTECH Summit 11000 Prober, as shown in Figure 3.

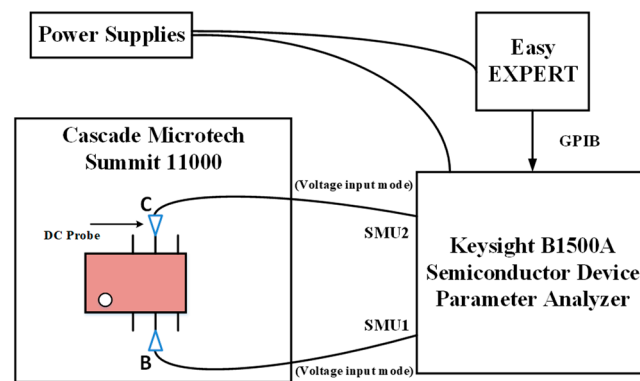


Figure 2. Diagram of electrical stress application.



Figure 3. Testing equipment.

3. Model Technique

3.1. Prior Knowledge Construction

For prior knowledge neural network-based aging modeling, the first step is that we need to obtain a priori aging knowledge in the form of physical/empirical formulas or equivalent circuits, which needs to represent the degradation tendency of the device characteristics of interest but not necessarily be very accurate or complete. In this paper, we complete the construction of the prior knowledge by modeling the HBT aging equivalent circuit.

In our previous study [19], the small-signal model with base-distributed capacitances applied to HBTs was proposed, as shown in Figure 4. The equivalent circuit model consists of three modules: parasitic elements, extrinsic distributed capacitance elements, and intrinsic model elements. The outermost part of the model is the parasitic module. C_{pbe} , C_{pce} , and C_{pbc} are the base-emitter, collector-emitter, and base-collector parasitic capacitances, respectively. L_b , L_c , and L_e are the lead inductances associated with the base, collector, and emitter, respectively. The above pad parasitic parameters are extracted by the Open and Short Test Structure method described in [20]. R_b , R_c , and R_e are the series resistances associated with the base, collector, and emitter, respectively, which can be determined in the cut-off condition measurements [21]. After peeling off the parasitic elements mentioned above, the periphery of the equivalent circuit features extrinsic distributed elements, in which C_{bcx} is the base-collector distributed capacitance and C_{bex} is the base-emitter distributed capacitance. Intrinsic model elements include the dynamic base resistance R_{bi} , the dynamic base-emitter resistance R_{be} , the intrinsic base-emitter capacitance C_{be} , the intrinsic base-collector capacitance C_{bc} , the DC transconductance G_{m0} , and the delay time τ . We use the peeling algorithm from our earlier work [19] to obtain the values of these extrinsic and intrinsic elements.

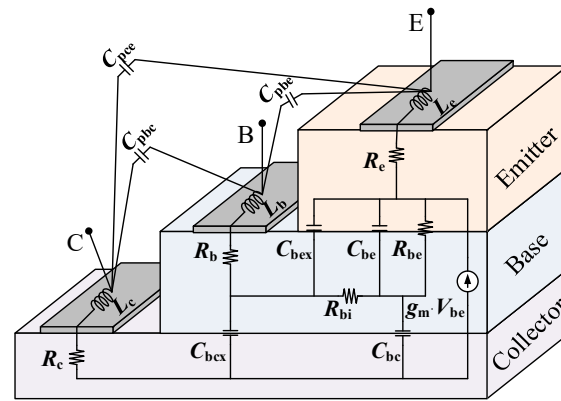


Figure 4. HBT small-signal equivalent circuit.

The aging reliability test data obtained from different batches of InP HBT devices under different stress times and stress magnitudes are characterized by the above-mentioned small-signal model topology and parameter extraction to determine the key model parameters affecting the characterization of degradation of HBT devices. Inserting a parameter ΔP related to the variation in the above key parameters due to aging effects, ΔP is defined as the difference between the key model parameters after and before degradation, which can be extracted experimentally by its variation with different stress time and stress magnitude conditions. Thus, the aging Equation (1) that can describe the degradation trend of the key parameters of the model is given. For the function establishment of the degradation amount parameter ΔP , we use a double exponential type function for modeling, as shown in Equation (2).

$$P_{\text{aging}}(V_{\text{CB, stress}}, t) = \Delta P(V_{\text{CB, stress}}, t) + P_{\text{initial}} \tag{1}$$

$$\Delta P(V_{\text{CB, stress}}, t) = A_0(1 - \exp((\exp(-\mu \cdot V_{\text{CB, stress}}) - 1) \cdot at)) \tag{2}$$

where P_{aging} is the value of key model parameters after degradation, ΔP is the amount of degradation of key model parameters, and P_{initial} is the value of key parameters in the fresh state. A_0 is the accelerated degradation saturation factor, a is the degradation acceleration factor, and μ is the degradation acceleration index factor. The degradation equations for the above key parameters are substituted back to the model topology shown in Figure 4 to obtain the degradation trend of the S -parameters of the device, and we use this rough degradation law as a priori degradation knowledge to improve the quality of machine learning-assisted aging modeling mapping.

3.2. Dual-Extreme Learning Machine(D-ELM) Structure

ELM neural network is usually a single-hidden layer feedforward neural network. It can randomly generate the weight between the input layer and the hidden layer and the bias of the neurons in the hidden layer, and there is no need to adjust them during the training process. Only the number of neurons in the hidden layer needs to be set to obtain the unique optimal solution [22].

The above traditional MLP structure for aging modeling is shown in Figure 5, which requires a large amount of degradation measurement data to construct the training set and more training time to learn the nonlinear relationships of the original problem to achieve the desired modeling accuracy. For the aging prediction of devices, therefore, we propose a D-ELM architecture, as shown in Figure 6, which decomposes the aging features and fresh features of devices into two sub-networks to simplify the complex nonlinear mapping of the original problem. NETWORK1 and NETWORK2 in Figure 6 have a common frequency input ($Freq$) and operating bias inputs (V_{CE}, I_b), and NETWORK2 has additional reliability parameter inputs ($V_{\text{CB, stress}}, t_{\text{stress}}$). Thus, we use NETWORK1 and NETWORK2 to characterize the fresh S -parameters and the degradation of S -parameters in HBTs, respectively. Then, the two sets of results are summed to obtain the aging S -parameters.

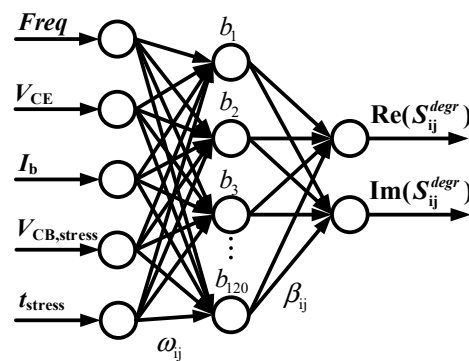


Figure 5. MLP structure for aging S-parameter modeling.

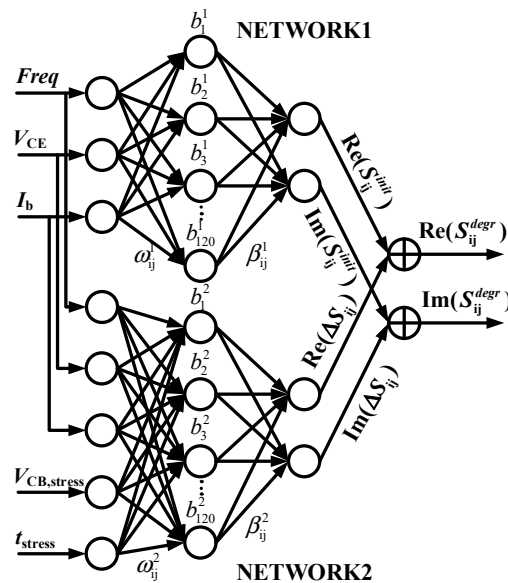


Figure 6. D-ELM neural network structure for aging S-parameter modeling.

In addition, we give a comparison of the time consumption and accuracy of the two structures in Table 1 with the same amount of data (1340 sample sets) and the percentage of the training set (75%) for the same problem. It is evident that the proposed D-ELM architecture has a much faster training speed and modeling accuracy; it saves 17.75% training time, and the accuracy is improved by about two times compared to the traditional MLP structure.

Table 1. Comparison of time consumption and accuracy of MLP and D-ELM structures.

Network Types	Time Required Per Training Session(s)	Residual Error (%)
MLP	11.72	8.22
D-ELM	9.64	3.93

3.3. PKNN-Based Aging Modeling Method

The D-ELM structure we proposed in Section 3.2 is a pure black-box model whose structure does not contain any relevant knowledge of the problem to be solved, and only the training data determine the structure of the network input–output mapping, so a large number of dataset sources are required to ensure the validity of the model.

Therefore, to link the prior knowledge constructed in Section 3.1, we further improve the D-ELM structure. For the degradation network (NETWORK2), we construct the degradation formulas of the key parameters and the aging equivalent circuit as the prior

knowledge module to be injected to obtain the prior knowledge-based degradation network (NETWORK2'). Subsequently, we employ this network in conjunction with NETWORK1 to simulate the S-parameters of the device in the fresh state to obtain the knowledge-based reliability neural network model, whose modeling framework is shown in Figure 7. The mathematical derivation process is shown in the following equations:

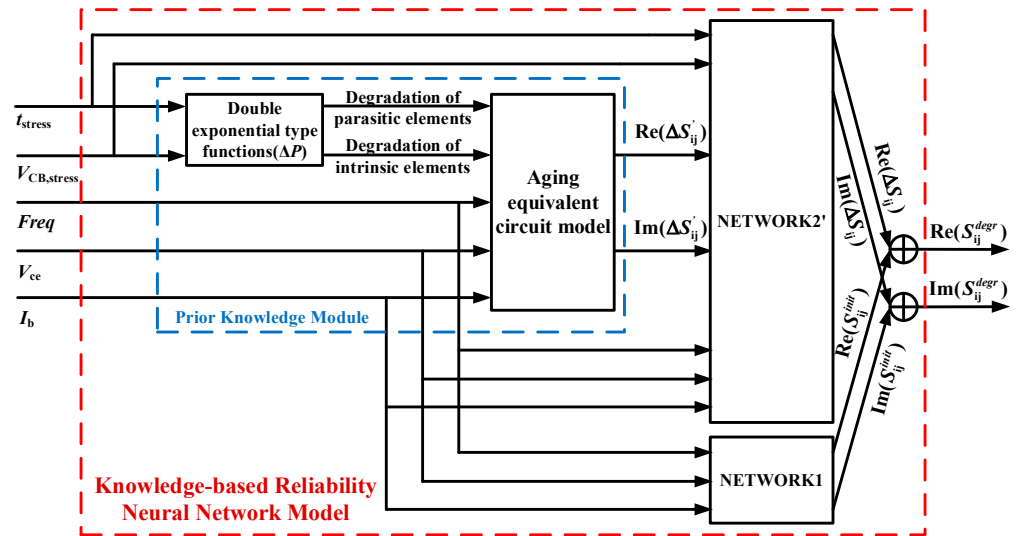


Figure 7. Knowledge-based reliability neural network modeling framework.

For NETWORK1 and NETWORK2', their input vectors are shown in Equations (3) and (4), respectively:

$$X_q^{N1} = [Freq_q, V_{ce,q}, I_{b,q}]^T, (q = 1, 2, \dots, T1) \tag{3}$$

$$X_q^{N2} = [Freq_q, V_{ce,q}, I_{b,q}, V_{CB,stress,q}, t_{stress,q}, Re(\Delta S'_{ij})_q, Im(\Delta S'_{ij})_q]^T, (q = 1, \dots, T2) \tag{4}$$

where N1 and N2 represent NETWORK1 and NETWORK2', respectively. T1 and T2 are the number of training set samples for NETWORK1 and NETWORK2', respectively. $\Delta S_{ij}'$ is the output of the coarse model degradation obtained from the aging equivalent circuit. The corresponding training target matrices of the two networks are shown in Equations (5) and (6) below:

$$T^{N1} = \begin{bmatrix} Re(S_{ij}^{init})_1 & Re(S_{ij}^{init})_2 & \dots & Re(S_{ij}^{init})_{T1} \\ Im(S_{ij}^{init})_1 & Im(S_{ij}^{init})_2 & \dots & Im(S_{ij}^{init})_{T1} \end{bmatrix}_{8 \times T1} \tag{5}$$

$$T^{N2} = \begin{bmatrix} Re(\Delta S_{ij})_1 & Re(\Delta S_{ij})_2 & \dots & Re(\Delta S_{ij})_{T2} \\ Im(\Delta S_{ij})_1 & Im(\Delta S_{ij})_2 & \dots & Im(\Delta S_{ij})_{T2} \end{bmatrix}_{8 \times T2} \tag{6}$$

where S_{ij}^{init} is the value of S-parameters in the fresh state. S_{ij}^{degr} is the value of S-parameters after degradation.

In addition, the connection weight vectors between the input and hidden layers of NETWORK1 and NETWORK2' are shown in Equations (7) and (8), and the bias vectors of the hidden neurons are shown in Equations (9) and (10). They can be randomly assigned before training and remain unchanged during the training process.

$$W_p^{N1} = [\omega_{p1}^{N1}, \omega_{p2}^{N1}, \omega_{p3}^{N1}], (p = 1, 2, \dots, H1) \tag{7}$$

$$W_p^{N2} = [\omega_{p1}^{N2}, \omega_{p2}^{N2}, \dots, \omega_{pH2}^{N2}], (p = 1, 2, \dots, H2) \tag{8}$$

$$B^{N1} = [b_1^{N1}, b_2^{N1}, \dots, b_{H1}^{N1}]^T \tag{9}$$

$$B^{N2} = [b_1^{N2}, b_2^{N2}, \dots, b_{H2}^{N2}]^T \tag{10}$$

where H1 and H2 represent the number of hidden layer neurons in NETWORK1 and NETWORK2', respectively. The transfer function of the hidden layer neuron is

$$g(\gamma) = \frac{1}{1 + e^{-\gamma}} \tag{11}$$

where γ is the weighted sum of all inputs to the neuron.

After all the functional variables in the neural network algorithm are defined, we can obtain the hidden layer output matrix as shown in Equations (12) and (13).

$$H^{N1} = \begin{bmatrix} g(W_1^{N1} X_1^{N1} + b_1^{N1}) & g(W_2^{N1} X_1^{N1} + b_2^{N1}) & \dots & g(W_{H1}^{N1} X_1^{N1} + b_{H1}^{N1}) \\ g(W_1^{N1} X_2^{N1} + b_1^{N1}) & g(W_2^{N1} X_2^{N1} + b_2^{N1}) & \dots & g(W_{H1}^{N1} X_2^{N1} + b_{H1}^{N1}) \\ \vdots & \vdots & \ddots & \vdots \\ g(W_1^{N1} X_{T1}^{N1} + b_1^{N1}) & g(W_2^{N1} X_{T1}^{N1} + b_2^{N1}) & \dots & g(W_{H1}^{N1} X_{T1}^{N1} + b_{H1}^{N1}) \end{bmatrix}_{T1 \times H1} \tag{12}$$

$$H^{N2} = \begin{bmatrix} g(W_1^{N2} X_1^{N2} + b_1^{N2}) & g(W_2^{N2} X_1^{N2} + b_2^{N2}) & \dots & g(W_{H2}^{N2} X_1^{N2} + b_{H2}^{N2}) \\ g(W_1^{N2} X_2^{N2} + b_1^{N2}) & g(W_2^{N2} X_2^{N2} + b_2^{N2}) & \dots & g(W_{H2}^{N2} X_2^{N2} + b_{H2}^{N2}) \\ \vdots & \vdots & \ddots & \vdots \\ g(W_1^{N2} X_{T2}^{N2} + b_1^{N2}) & g(W_2^{N2} X_{T2}^{N2} + b_2^{N2}) & \dots & g(W_{H2}^{N2} X_{T2}^{N2} + b_{H2}^{N2}) \end{bmatrix}_{T2 \times H2} \tag{13}$$

In turn, the output matrices of the two networks are obtained separately:

$$\hat{T}^{N1} = f(Freq, V_{ce}, I_b) = H^{N1} \beta^{N1} \tag{14}$$

$$\hat{T}^{N2} = f(Freq, V_{ce}, I_b, V_{CB, stress}, t_{stress}, Re(\Delta S'_{ij}), Im(\Delta S'_{ij})) = H^{N2} \beta^{N2} \tag{15}$$

where β^{N1} and β^{N2} are the connection weight matrices between the hidden and output layers to be solved during the neural network training, respectively. The optimal weight between the hidden and output layers can be found by obtaining the Moore–Penrose generalized inverse [23] of the output matrix of the hidden layer to complete the training of the ELM network. In summary, we can obtain the objective function of the proposed PKNN-based D-ELM network training as shown in Equation (16), and its least squares solution can be obtained from Equation (17).

$$\min(\|\hat{T}^m - T^m\|^2) = \min(\|H^m \beta^m - T^m\|^2), (m = N1 \text{ or } N2) \tag{16}$$

with

$$\hat{\beta}^m = (H^m)^\dagger T^m \tag{17}$$

At this point, the network structure, corresponding neuron weights, and bias of the networks are determined. In addition to the reliability parametric inputs of the original problem, the S-parameter degradation information obtained from the aging small-signal equivalent circuit is used as an additional input to NETWORK2' in the constructed knowledge-based reliability neural network model. Then, the input–output mapping to be learned by the neural network is the mapping between the output of the existing

approximate reliability model and the original problem, which provides the degradation knowledge to weaken the strong nonlinear relationship between the reliability coefficients and the degradation of *S*-parameters and improve the quality of the mapping. This, in turn, serves to reduce the training dataset requirement and improve the network generalization capability and modeling efficiency.

3.4. Model Optimization

For traditional ELM networks, random selectivity makes the accuracy and convergence of neural networks relatively unstable. Typically, a network needs to be initialized many times to try out the optimal value. The adaptive genetic algorithm (AGA) is an evolutionary algorithm, a method to search for the optimal solution by simulating the natural evolutionary process [24]. Therefore, we use the AGA to automatically optimize the input weights and hidden layer bias of the networks to avoid the effects of random initialization.

This optimization process is described in the flowchart shown in Figure 8. First, we need to prepare the training sample sets of input and output data with measured fresh/degraded *S*-parameters for NETWORK1 and NETWORK2'. After normalizing the above training data, the connection weight between the input and hidden layers and the bias of the neurons in the hidden layer are randomly initialized in the D-ELM network shown in Figure 6. Since the AGA cannot directly deal with the parameters of the problem space, they need to be represented as individuals in the genetic space by coding. A fitness evaluation is then performed for each individual. The fitness function is expressed by Equation (18).

$$fitness = 1 / \sum_{i=1}^N |y_{mea} - y_{pre}| \tag{18}$$

where y_{mea} and y_{pre} represent the measured output and the predicted output of the neural network, respectively. N is the total number of training data.

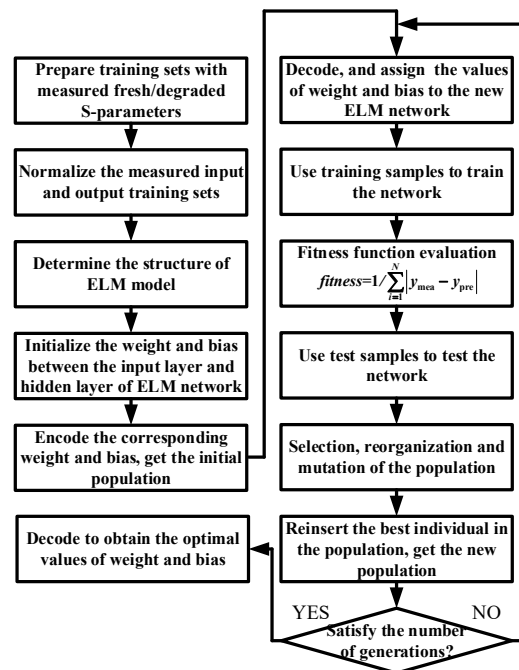


Figure 8. Flowchart of AGA optimization.

After the fitness evaluation, 15% of individuals in the paternal generation will be rejected, and the remaining individuals will undergo reorganization and mutation to obtain offspring. The corresponding reorganization probabilities p_x and mutation probabilities p_m are set to be adaptive, as shown in Equations (19) and (20):

$$p_x = \begin{cases} k_1(f_{\max} - f) / (f_{\max} - f_{\text{avg}}), & f \geq f_{\text{avg}} \\ k_2, & f < f_{\text{avg}} \end{cases} \quad (19)$$

$$p_m = \begin{cases} k_3(f_{\max} - f') / (f_{\max} - f_{\text{avg}}), & f' \geq f_{\text{avg}} \\ k_4, & f' < f_{\text{avg}} \end{cases} \quad (20)$$

where f_{\max} is the maximum fitness in the population. f_{avg} is the average fitness of the population. $f_{\max} - f_{\text{avg}}$ reflects the convergence state of optimization. f is the larger fitness of the two individuals to be reorganized. f' is the fitness of the individual to be mutated. k_1 , k_2 , k_3 , and k_4 are constants less than 1.

For high-quality individuals with high fitness, p_x and p_m are reduced to preserve them; for low-quality individuals with low fitness, p_x and p_m are increased to eliminate them. In addition, in the early stage of iterative optimization, the population needs a large reorganization and mutation probability to achieve a fast search for the optimal solution, while in the late stage of convergence, the population needs a small reorganization and mutation probability to enable the population to converge quickly after searching for the optimal solution. Then, the superior individuals in the offspring were reinserted into the paternal generation to form a new population. The above process will not stop until the number of training cycles satisfies the number of evolutionary generations set in the experiment. Finally, we decode to obtain the optimal values of the weight between the input layer and the hidden layer and the bias of the hidden neurons. At this point, we can obtain the trained AGA-D-ELM neural network model to predict the aging S -parameters.

4. Results and Discussion

In order to more intuitively represent the modeling process in this paper, we use the PKNN-based aging small-signal modeling flowchart shown in Figure 9 to briefly summarize the above steps of device small-signal aging knowledge construction, knowledge injection, D-ELM network training, and network optimization. In the first stage, we use NETWORK1 to train and predict the S -parameter samples in the fresh state of the device. Then, we transform the degradation trend of the key parameters into the degradation information of the S -parameters after applying stress to the device through the aging compact model in the second stage. In the third stage, we inject the rough aging information obtained in the second stage into the degradation NETWORK2, then train and optimize this network. Finally, the aging S -parameters of the device are obtained by accumulating the predicted outputs of the two networks under the corresponding degradation conditions. The proposed aging modeling technique is evaluated and validated by modeling the aging S -parameters of different batches of InP HBTs degraded under different stress magnitudes and times in the frequency range of 0.1–40 GHz.

Trained artificial neural network models are usually evaluated using the prediction error of the test sample sets. For the reliability of the ANN model in this paper, the main factors affecting its network training in terms of dataset sample requirements are the total reliability sample size and the percentage of the training sample sets. Therefore, we set up seven groups of training experiments, as summarized in Table 2, to verify the validity and accuracy of the proposed model. The first, second, and third groups keep the same total sample size (2412 sets) and 50% training set share. These three experiments were conducted to model the degradation of each component of the S -parameters using the traditional ELM structure and the proposed PKNN structure, respectively, to compare the degradation prediction ability of the two networks under the same sample size and the same training conditions with different amounts of knowledge injection. Additionally, to further illustrate

the superiority of the aging modeling approach proposed in the paper, keeping the same conditions as the first three sets of experiments, the widely used backpropagation (BP) and radial basis function (RBF) algorithms are applied for additional comparison in the fourth and fifth groups, respectively. Then, we set up a sixth set of experiments, still keeping the same total sample size (2412 sets) as the first five sets of experiments but randomly selecting 80% of the total sample size to train the network on the traditional ELM structure. This was conducted to compare the results with the second and third sets of experiments on prediction with the proposed PKNN structure to verify the prediction ability of the proposed model after the dependence on the reliability training samples is reduced. Finally, we supplemented a seventh set of experiments by taking about 60% (1458 sets) of the total sample size of the first six sets, with the training set share remaining the same as that of the sixth set. Then, we compared it with the prediction results of the ELM structure in the sixth set of experiments to further compare and illustrate the prediction ability of the proposed PKNN model with the reduction in the total reliability sample size. In addition, after keeping the same total sample size and fixing the training set share for the third and sixth sets of experiments, the test sets they used for validation also naturally formed a set of control groups with a share of 50% and 20%, respectively. The results of the comparison between these two sets of experiments can be used to further illustrate how well the generalization ability of both the traditional ELM structure and the proposed PKNN structure is. In Table 3, we give the mean relative errors of the prediction of the degradation of each component of the S-parameters in the seven sets of experimental test sets, and their error equations are shown below.

$$MRE = \frac{100\%}{N} \cdot \sum_{i=1}^N \left| \frac{y_{mea} - y_{pre}}{y_{mea}} \right| \tag{21}$$

where y_{mea} and y_{pre} represent the measured output and the predicted output of the neural network, respectively. N is the number of data.

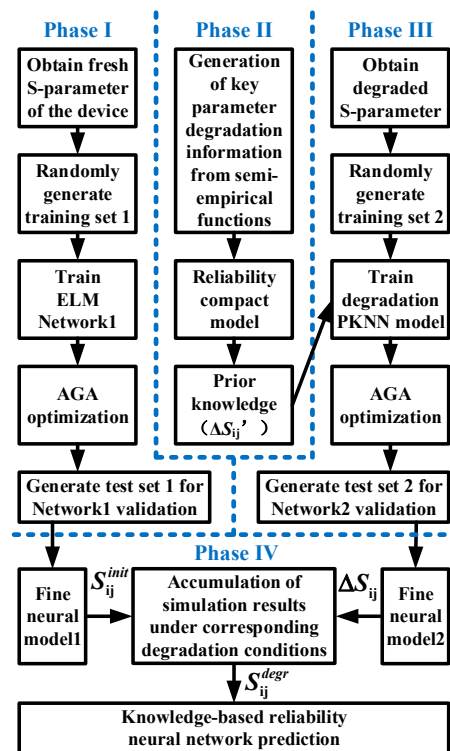


Figure 9. Flowchart of the proposed PKNN-based aging modeling method.

Table 2. Experimental setup of the dataset for training and validation.

Group	Training Networks	Knowledge Samples	Knowledge Injection Share (%)	Total Sample Size	Training Data Share (%)	Training Sample Sets	Test Sample Sets
1	ELM	0	0	2412	50	1206	1206
2	PKNN	4824	50	2412	50	1206	1206
3	PKNN	9648	100	2412	50	1206	1206
4	BP	0	0	2412	50	1206	1206
5	RBF	0	0	2412	50	1206	1206
6	ELM	0	0	2412	80	1930	482
7	PKNN	9328	100	1458	80	1166	292

Table 3. Comparison of prediction accuracy for seven groups of training experimental settings.

Group	MRE for Test Sets (%)								Total Error (%)
	$\Delta\text{Re}(S11)$	$\Delta\text{Im}(S11)$	$\Delta\text{Re}(S12)$	$\Delta\text{Im}(S12)$	$\Delta\text{Re}(S21)$	$\Delta\text{Im}(S21)$	$\Delta\text{Re}(S22)$	$\Delta\text{Im}(S22)$	
1	1.27	2.88	1.22	3.84	4.17	4.01	4.25	5.13	3.35
2	0.96	1.39	1.12	2.98	1.82	1.37	2.59	1.97	1.78
3	0.16	0.62	0.38	1.86	1.28	0.40	1.10	1.04	0.86
4	2.59	8.07	9.69	13.36	0.95	1.31	3.84	2.25	5.26
5	2.55	3.27	10.43	19.52	0.23	1.17	1.96	6.24	5.67
6	1.04	1.40	0.76	2.73	1.94	2.29	3.72	2.38	2.01
7	0.51	0.38	0.32	1.47	0.67	0.47	0.98	1.09	0.74

From the comparison of the prediction results between the first and third sets of experiments, the overall prediction accuracy of the PKNN model for the degradation of each component of the S -parameter is better than that of the ELM structure. The improvement is at least 51.6% for each component and a reduction of 74.3% in the overall error for the same sample size and training configuration. For the fourth and fifth groups of BP and RBF algorithm networks, the overall prediction accuracy of PKNN is about three times higher than theirs. From the comparison of the prediction results between the third and sixth sets of experiments, compared with the aging prediction of the ELM structure without knowledge injection, the PKNN-based aging model can meet or even exceed its prediction accuracy for all the components of the S -parameters with at least a 37.5% reduction in the amount of training data. Furthermore, for the second set of experiments with only 50% knowledge injection, although the accuracy of the PKNN model decreased compared to the third set, the total prediction accuracy still improved by 11.4% compared to the sixth set. Moreover, the size of the test samples to be validated for the PKNN model in the third set of experiments (1206 sets) is about 2.5 times larger than that of the ELM model in the sixth set of experiments (482 sets), while its overall error has been reduced by 57.2%, which is evident in the significant improvement of its generalization ability. From the comparison of the prediction results between the sixth and seventh sets of experiments, the proposed PKNN model reduces its overall error by 63.2%, with the total reliability sample size reduced by at least 39.6% while controlling the same occupancy of the training set, further demonstrating the superiority of the proposed model in terms of reliability sample dependency while guaranteeing accuracy.

In addition, we have further compared the third and sixth sets of experiments, which have been verified to have improved prediction accuracy and generalization ability in terms of the efficiency and convergence of the network training optimization. In this paper, for each set of training samples, the input layer of NETWORK1, which describes

the S -parameters of the fresh state of the device, has three input neurons, and the input layers of NETWORK2 and NETWORK2', which describe the amount of degradation of the S -parameters of the device, have five and thirteen input neurons, respectively. The D-ELM structure is a single hidden layer structure with 120 hidden neurons. Therefore, the number of individuals to be optimized in the AGA optimization process is 480 for the population of NETWORK1 and 720 and 1680 for the population of NETWORK2 and NETWORK2', respectively. The stopping criterion is set to a maximum of 1000 generations. The fitness evolution curve of NETWORK1 during the training iterations is shown in Figure 10a. The total optimization time is 1956, and its fitness value is optimized by 17.2%. For NETWORK2 using the ELM structure and NETWORK2' using the PKNN structure, their fitness evolution curves are shown in Figure 10b,c. The optimized convergence duration is 4939s and 2173s, respectively, and the convergence speed is accelerated by about 56.0%, and their fitness values are optimized by 34.6% and 30.8%, respectively. The results of the above networks in terms of efficiency and convergence of training and optimization are summarized in Table 4, from which it can also be seen that the overall fitness range of the PKNN aging model for its population during optimization iterations is about three times higher than that of the traditional ELM network. The high superiority of individuals throughout the optimization period further illustrates the advantages of the proposed model in terms of simulation accuracy.

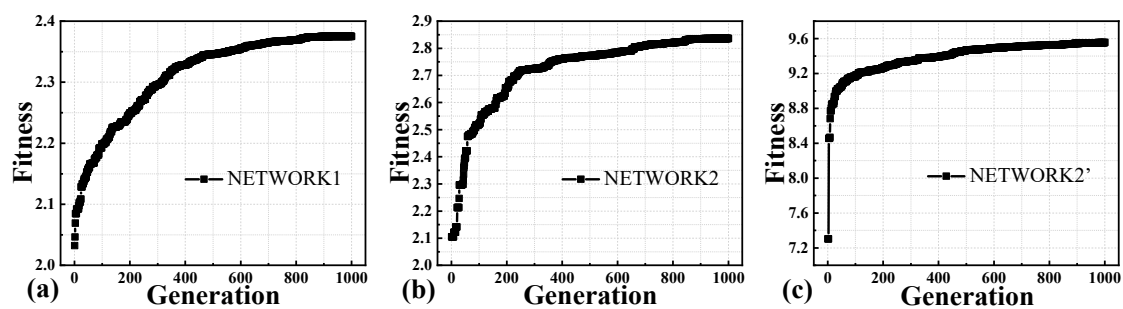


Figure 10. Evolution of fitness during optimization for (a) NETWORK1, (b) NETWORK2, and (c) NETWORK2'.

Table 4. Comparison of optimization efficiency and convergence of the networks.

Comparison	NETWORK1	NETWORK2	NETWORK2'
Fitness range	2.03~2.38	2.11~2.84	7.30~9.55
Total duration of optimization (s)	1956	5845	3728
Number of iterations to reach convergence	864	845	583
Convergence time (s)	1690	4939	2173

Using the prior knowledge-based AGA-D-ELM network developed above, we predict the S -parameters of InP DHBT devices in the fresh state and the degradation of the S -parameters under different stress magnitudes and stress times. A comparison of the simulations of NETWORK1 and the conventional MLP network for the fresh S -parameters is given in Figure 11, which further illustrates the accuracy of the predictions of the D-ELM structure. As shown in Figures 12a and 13a, we observed the performance of different batches of $0.7 \times 13 \mu\text{m}^2$ InP DHBT devices under the bias of ($V_{CE} = 1.6 \text{ V}$, $I_b = 200 \mu\text{A}$) and subjected to two stress conditions: ($V_{CB, \text{stress}} = 3.4 \text{ V}$, $t_{\text{stress}} = 250 \text{ min}$) and ($V_{CB, \text{stress}} = 4.3 \text{ V}$, $t_{\text{stress}} = 310 \text{ min}$). The model predictions of the degradation of the real and imaginary parts of each component of S -parameters are compared with the corresponding test results using the NETWORK2' and NETWORK2 trained in the second and third sets of experiments,

respectively. In addition, to further complement the predictive effectiveness of the proposed model, we present the simulation results of the model in Figures 12b and 13b for the magnitude and phase of the degradation of each S -parameter component in the Advanced Design System (ADS). The simulation accuracy of NETWORK2' on the degradation of S -parameters is still able to reach or even be better than that of NETWORK2 after the dependence on reliability training samples is reduced. It can be seen that the proposed PKNN network structure is able to predict the degradation trend of the S -parameters of the HBT devices after the accelerated aging well, which provides guidance for the establishment of a highly efficient and accurate aging model for microwave devices in practical applications.

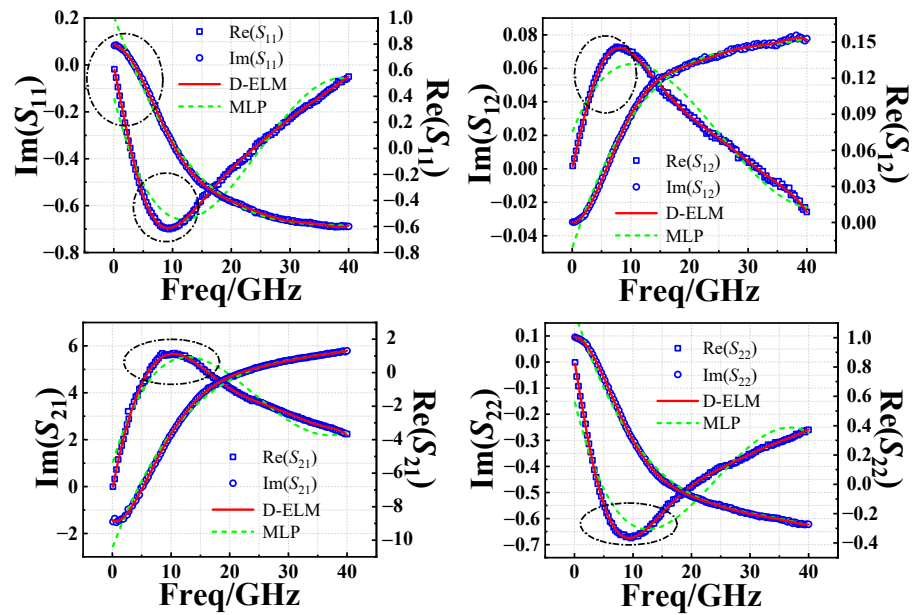


Figure 11. Comparison between predicted and measured results of the S -parameters in the fresh state from 0.1 to 40 GHz.

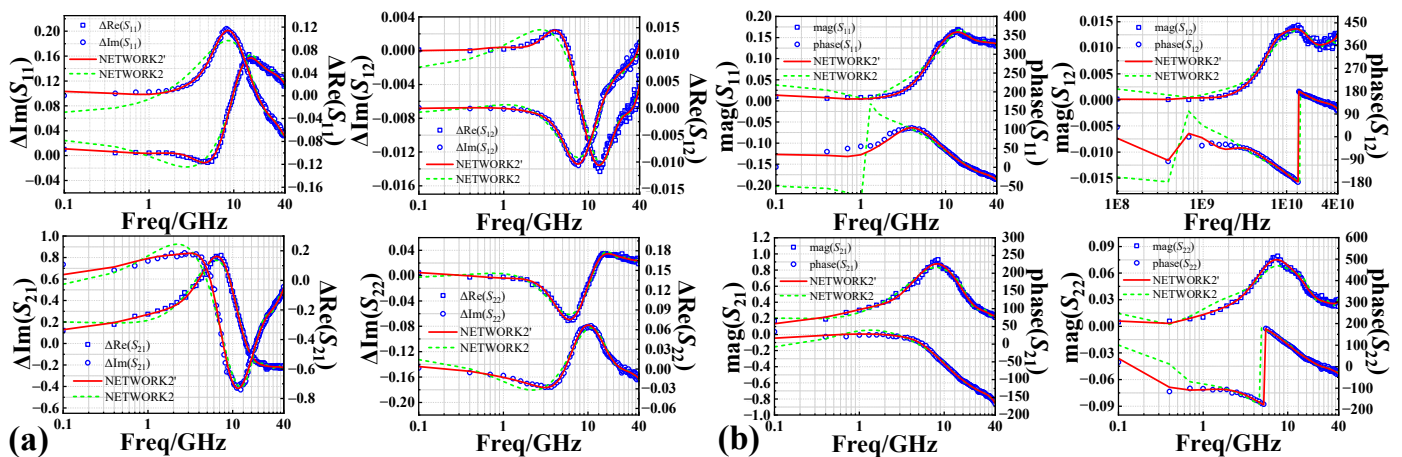


Figure 12. Comparison between predicted and measured results of (a) real and imaginary parts and (b) magnitude and phase of degradation of S -parameters from 0.1 to 40 GHz at $V_{CB, stress} = 3.4$ V, $t_{stress} = 250$ min.

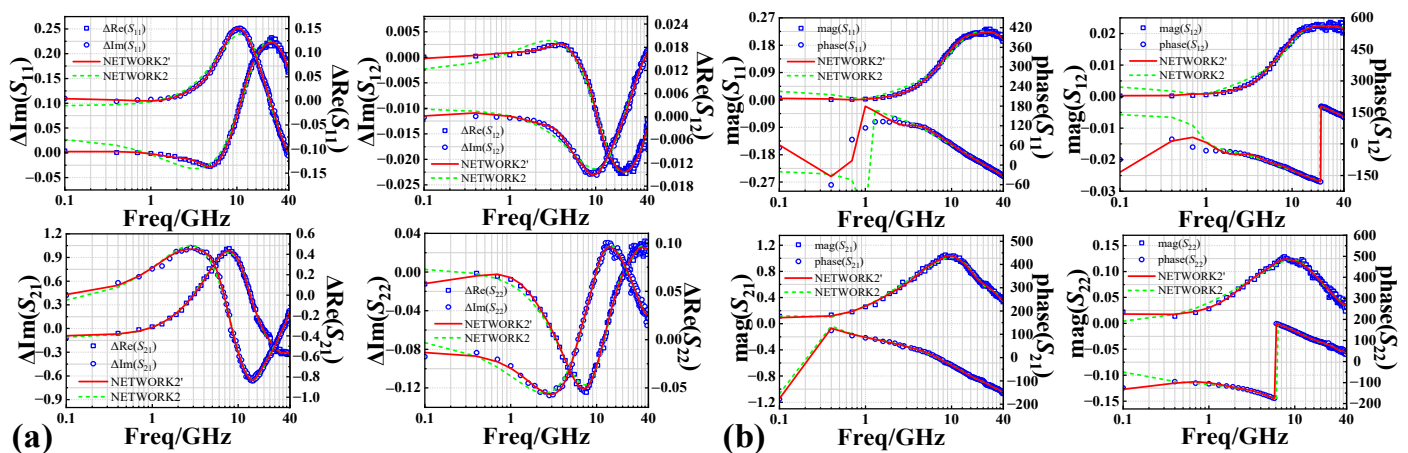


Figure 13. Comparison between predicted and measured results of (a) real and imaginary parts and (b) magnitude and phase of degradation of S-parameters from 0.1 to 40 GHz at $V_{CB, stress} = 4.3$ V, $t_{stress} = 310$ min.

5. Conclusions

A prior knowledge-based aging small-signal neural network model is proposed for the degradation prediction of S-parameters of microwave HBTs. The adopted dual-extreme learning machine architecture decomposes the aging features and the fresh features of the device into two sub-networks, respectively, to simplify the complex nonlinear mapping of the original problem. Moreover, we transform the degradation trend of the device key parameters into the degradation information of the S-parameters after the device is stressed by the aging compact model. Compared with the traditional MLP degradation sub-network architecture, the proposed knowledge-based reliability neural network model incorporates all this degradation information. This integration improves prediction accuracy with reduced reliability training sample dependency, as well as improving the generalization ability of the network. The superiority of the convergence efficiency of the proposed model in the optimization process is further verified by introducing an adaptive genetic algorithm to automatically optimize the input weights and hidden layer biases in the network to circumvent the instability caused by random initialization. Finally, the validity and accuracy of the proposed PKNN-based aging model are verified by comparing the predicted and measured results of the S-parameter degradation of different batches of $0.7 \times 13 \mu\text{m}^2$ InGaAs/InP DHBTs in the frequency range of 0.1–40 GHz.

Author Contributions: Conceptualization, L.C.; formal analysis, L.C. and J.Q. (Jiantao Qiao); funding acquisition, H.L., W.C., Y.Z. (Yimen Zhang) and Y.Z. (Yuming Zhang); investigation, L.C. and J.Q. (Junjun Qi); methodology, L.C. and S.Y.; project administration, H.L.; supervision, H.L. and C.L.; writing—original draft preparation, L.C.; writing—review and editing, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Defense Science and Technology Foundation Strengthening Program (Program No. 2019-XXXX-XX-236-00).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cheron, J.; Jones, R.D.; Williams, D.F.; Aaen, P.H. A Tunable 220 GHz Comb Generator Realized With an Ultrawideband Mixer in an InP HBT Technology. *IEEE Microw. Wirel. Technol. Lett.* **2023**, *33*, 1345–1348. [\[CrossRef\]](#)
2. Michinori, H.; Takashi, S.; Ryota, I.; Toshiaki, N. Millimeter-Wave Single-Pixel Imaging Using Electrically-Switchable Liquid-Crystal Mask. *IEICE Trans. Electron.* **2023**, *E106C*, 34–40.
3. Oh, M.-H.; Kwon, M.-W.; Park, K.; Park, B.-G. Sensitivity Analysis Based on Neural Network for Optimizing Device Characteristics. *IEEE Electron. Device Lett.* **2020**, *41*, 1548–1551. [\[CrossRef\]](#)

4. Choe, G.; Kwak, J.; Yu, S. Machine Learning-Assisted Compact Modeling of W-Doped Indium Oxide Channel Transistor for Back-End-of-Line Applications. *IEEE Trans. Electron. Devices* **2023**, 1–8. [[CrossRef](#)]
5. Chen, D.; Yang, Y.; Wang, X.; Li, D.; Liang, Y.; Xu, C. Thermal-Stress Coupling Optimization for Coaxial through Silicon Via. *Symmetry* **2023**, *15*, 264. [[CrossRef](#)]
6. Zatout, B.; Maafri, D.; Taibi, A.; Belaroussi, Y.; Kerai, S.; Al Sabbagh, M.; Yagoub, M.C.E. An efficient extrinsic capacitances extraction method for small-signal GaN HEMT devices. *Int. J. Numer. Model.-Electron. Netw. Devices Fields* **2023**, *36*, e3117. [[CrossRef](#)]
7. Mavredakis, N.; Pacheco-Sanchez, A.; Wei, W.; Pallecchi, E.; Happy, H.; Jiménez, D. Straightforward bias- and frequency-dependent small-signal model extraction for single-layer graphene FETs. *Microelectron. J.* **2023**, *133*, 105715. [[CrossRef](#)]
8. Wang, Y.; Ding, W.; Su, Y.; Yang, F.; Ding, J.; Zhou, F.; Jin, Z. An electromagnetic simulation assisted small signal modeling method for InP double-heterojunction bipolar transistors. *Chin. Phys. B* **2022**, *31*, 068502. [[CrossRef](#)]
9. Cai, M.; Leang, S.E.; Chew, K.W.; Tan, P.Y.; Herlambang, A.P.; Zhu, C.; Guo, Y. Analysis and Modeling for Reverse Body Bias Stress Impact on HCI Induced Degradation in n-Type EDMOS. In Proceedings of the 2022 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA), Singapore, 18–21 July 2022.
10. Kamrani, H.; Jabs, D.; D’Alessandro, V.; Rinaldi, N.; Jacquet, T.; Maneux, C.; Zimmer, T.; Aufinger, K.; Jungemann, C. Microscopic Hot-Carrier Degradation Modeling of SiGe HBTs Under Stress Conditions Close to the SOA Limit. *IEEE Trans. Electron. Devices* **2017**, *64*, 923–929. [[CrossRef](#)]
11. Yamaguchi, Y.; Oishi, T. Quasi-Physical Equivalent Circuit Model of RF Leakage Current in Substrate Including Temperature Dependence for GaN-HEMT on Si. *IEEE Trans. Microw. Theory Tech.* **2022**, *71*, 1945–1956. [[CrossRef](#)]
12. Sun, Y.-B.; Fu, J.; Wang, Y.-D.; Zhou, W.; Zhang, W.; Liu, Z.-H. Extraction of temperature dependences of small-signal model parameters in SiGe HBT HICUM model. *Chin. Phys. B* **2016**, *25*, 048501. [[CrossRef](#)]
13. Kabir, H.; Zhang, L.; Yu, M.; Aaen, P.H.; Wood, J.; Zhang, Q.-J. Smart Modeling of Microwave Devices. *IEEE Microw. Mag.* **2011**, *11*, 105–118. [[CrossRef](#)]
14. Jarndal, A. Neural network electrothermal modeling approach for microwave active devices. *Int. J. RF Microw. Comput.-Aided Eng.* **2019**, *29*, e21764. [[CrossRef](#)]
15. Jacquet, T.; Sasso, G.; Chakravorty, A.; Rinaldi, N.; Aufinger, K.; Zimmer, T.; d’Alessandro, V.; Maneux, C. Reliability of high-speed SiGe:C HBT under electrical stress close to the SOA limit. *Microelectron. Reliab.* **2015**, *55*, 1433–1437. [[CrossRef](#)]
16. Wang, H.; Ng, G.I. Investigation of the degradation of InGaAs/InP double HBTs under reverse base-collector bias stress. *IEEE Trans. Electron. Devices* **2001**, *48*, 2647–2654. [[CrossRef](#)]
17. Rehnmark, S. On the calibration process of automatic network analyzer system. *IEEE Trans. Microw. Theory Tech.* **1974**, *22*, 457–458. [[CrossRef](#)]
18. Ferrero, A.P.; Pisani, U. Twoport network analyzer calibration using an unknown “Thru”. *IEEE Microw. Guid. Wave Lett.* **1992**, *2*, 505–507. [[CrossRef](#)]
19. Cheng, L.; Lu, H.; Xia, M.; Cheng, W.; Zhang, Y.; Zhang, Y. An augmented small-signal model of InP HBT with its analytical-based parameter extraction technique. *Microelectron. J.* **2022**, *121*, 105366. [[CrossRef](#)]
20. Gao, J.J. *Heterojunction Bipolar Transistors for Circuit Design: Microwave Modeling and Parameter Extraction*; Wiley: Singapore, 2015.
21. Bousnina, S.; Falt, C.; Mandeville, P.; Kouki, A.B.; Ghannouchi, F.M. An accurate on-wafer de-embedding technique with application to HBT devices characterization. *IEEE Trans. Microw. Theory Tech.* **2002**, *50*, 420–424. [[CrossRef](#)]
22. Huang, G.; Huang, G.B.; Song, S.J.; You, K.Y. Trends in extreme learning machines: A review. *Neural Netw.* **2015**, *61*, 32–48. [[CrossRef](#)]
23. Ning, K.; Liu, M.; Dong, M.; Wu, C.; Wu, Z. Two Efficient Twin ELM Methods With Prediction Interval. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*. [[CrossRef](#)] [[PubMed](#)]
24. Locatelli, M.; Schoen, F. (Eds.); *Global Optimization: Theory, Algorithms and Applications*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2013.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.