A Systematic Approach to a Reliable Neural Model for pHEMT Using Different Numbers of Training Data

Mojtaba Joodaki and Günter Kompa

Dept. of High Frequency Engineering, University of Kassel, Kassel, D-34121, Germany

***Abstract* — A systematic approach is presented to achieve a reliable neural model for microwave active devices with different numbers of training data. The method is implemented for a small-signal bias depended modeling of pHEMT with different numbers of training data. The errors for different numbers of training data have been compared to each other and show that by using this method a reliable model is achievable even though the number of training data is considerably small. The method aims at constructing a model which can satisfy the criteria of minimum training error, maximum smoothness (to avoid the problem of over-fitting), and simplest network structure.**

I. INTRODUCTION

Neural networks recently have been used as fast and flexible tools for microwave modeling, simulation, and optimization. Since they are fast, accurate, flexible, and can be constructed from microwave data under different conditions, they are an excellent solution for RF and microwave design considering packaging effects.

Despite the outstanding advantages of neural networks, as the number of model input parameters increase, the amount of training data, size of neural network, and training time would all increase. Increased nonlinearity in a model also requires increased training data, larger neural network size, and longer training time [1]. In this way decreasing the number of the training data is of utmost importance.

On the other hand, design of experiment (DOE) can be used to obtain maximum information with minimal measured data and produce an experimental plan that is in some sense mathematically and statistically optimal. Watson and Gupta used DOE to decrease the number of electromagnetic simulations in electromagnetically trained artificial neural networks (EM-ANN) models for microstrip vias and interconnects in multilayer circuits [2].

A very important issue is, when ample training data are available for model development involving ANN, considerable freedom exists in the selection of model topology. However, when only limited data are available, the size of a network and the number of connections have to be carefully selected, and the network should be appropriately trained; otherwise, it will not generalize well even when interpolating. In this paper, to have a reliable neural model from a limited number of training data, DOE method is used to prepare the proper input training data to the neural network and then the neural model is constructed with respect to minimum training error, maximum smoothness, and simplest network structure. The method is implemented for constructing a small-signal bias-dependent neural model for a pHEMT from a limited number of training data.

A neural network HBT modeling technique had been already introduced by Devadhaktuni, Xi, and Zhang [3]. They implemented a multilayer perceptrons (MLP) neural network with 3 layers to model the HBT directly from its S-parameter data under different bias conditions. A similar model to [3] has been used in this paper and the only differences are in inputs and training algorithm. The MLP type neural network in this paper takes three inputs (frequency, drain-source voltage, and gate-source voltage) and gives 8 outputs (magnitudes and phases of S_{11}, S_{12}, S_{21}, and S_{22}). The most commonly used training algorithm of Back-Propagation, considering a smoothing term for adjusting the weights, has been used. Predictions of the neural model is tested by a new set of measured data. The results with different number of training data by using this method have been compared with those of a neural model trained with a large number of training data.

II. METHOD FOR TRAINING AND EVALUATION OF THE NEURAL MODEL

Hung and et al. [4] proposed a criterion, tempered modified prediction squared error (MPSE), that can be used to train and evaluate a neural model. The criterion for a single output is

$$MPSE = TSE + \alpha^2 \sigma_0^2 \frac{k}{N} \tag{1}$$

where

$$TSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{2}$$

is the total training squared error, α is the penalty factor in the range of 0 to 1, k is the number of coefficients in a

0-7803-7239-5/02/$10.00 © 2002 IEEE

network, y_i is the ith measurement which is the ith target of the output node (i = 1, 2, ..., N), $\hat{y}_i$ is the ith output of output node and the error variance of σ_0^2 is

$$\sigma_0^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N} \quad (3)$$

where $\bar{y}$ is the mean value of all the experimental data.

Because, according to Kolmogorov's theorem [5], any continuous vector mapping of a vector varible on any compact (closed and bounded) set can be implemented exactly with a three-layered artificial neural network three-layered neural model has been used. With the full connections between each pair of adjacent layers, the total number of weights in a three-layered neural network, k, is express as

$$k = (N_i \times N_o)N_h \quad (4)$$

where N_i is the number of input neurons, N_o is the number of output neurons and N_h is the number of hidden neurons. Thus for a multi output neurons the *MPSE* is

$$MPSE = \frac{1}{N}\sum_{j=1}^{N_o}\sum_{i=1}^{N}(y_{i,j} - \hat{y}_{i,j})^2$$

$$+\alpha^4 \frac{(N_i + N_o)N_h}{N^2}\sum_{j=1}^{N_o}\sum_{i=1}^{N}(y_{i,j} - \frac{\sum_{i=1}^{N} y_{i,j}}{N}) \quad (5)$$

where $y_{i,j}$ is the jth observed output of the ith experiment and $\hat{y}_{i,j}$ is the jth estimated output of the ith experiment.

By increasing the number of hidden neurons, TSE decreases exponentially but the second part of (5) increases linearly. Because the MPSE is the sum of the both parts, it decreases initially, but eventually the MPSE starts to increase due to the rise in the second part. The location of minimum MPSE is function of α. In this work α was selected between 0.4 and 0.6, so that the value of $MPSE_{min}$ deceased neither rapidly nor very slowly and over-fitting is avoided while the best network was reasonably complicated.

To add another smoothing factor, the modified delta learning rule proposed by Sejnowski and Rosenberg [5] has been used to provide exponential smoothing in weight adjustment. The change of weights between adjacent layers is expressed as

$$\Delta w_{pq,l}(n+1) = (1-\beta_l)\delta_{q,l}O_{p,l} + \beta_l \Delta w_{pq,l}(n) \quad (6)$$

and the adjustment of the weights is done in the following manner:

$$w_{pq,l}(n+1) = w_{pq,l}(n) + \eta_l \Delta w_{pq,l}(n+1) \quad (7)$$

where $w_{pq,l}(n)$ is the weight from neuron p in layer l to neuron q in layer (l+1) at step n (before adjustment); $w_{pq,l}(n+1)$ is the weight from neuron p in layer l to neuron q in layer (l+1) at step (n+1) (after adjustment); $\delta_{q,l}$ is the error term of neuron q between target and actual outputs of layer l; $O_{p,l}$ is the actual output of neuron p in layer l; β_l is the smoothing factor in layer l; η_l is the learning rate in layer l.

The range of β_l is from 0 to 1. If β_l is 0, then smoothing is minimum; the entire weight adjustment comes from the newly calculated change. If β_l is 1, the new adjustment is ignored and pervious one is repeated.

An interactive approach has been used to train and evaluate a neural model as follow:

1. Select the number of training data.
2. Determine the minimum and maximum numbers of hidden layer neurons to be considered. N_t is the total number of neural network to be tested.
3. Construct the neural networks with the minimum number of neurons in the hidden layer up to maximum number of neurons in the hidden layer. Input nodes are V_{ds}, V_{gs} and frequency and a bias neuron (4 nodes). The output are magnitudes and phases of S-parameters (8 nodes).
4. Train the neural networks considering (6) and (7).
5. Calculate the TSE, $\alpha^4\hat{\sigma}_0^2 k/N$, and MPSE for different networks. Plot MPSE as a function of hidden neurons for different α with discrete intervals of 0.1 ranging from 0 to 1. Plot TSE and $\alpha^4\hat{\sigma}_0^2 k/N$ as a function of hidden layer neurons.
6. Select a neural model using two figures obtained in step 5 or use the default α between 0.3 to 0.7.

III. Model Development and Results

To verify the method, crossed D-optimal design has been used to design the bias dependent S-parameters measurement. The levels for Vds are 0 (0.1) 1 V and 1.5 (0.5) 5 V and for Vgs are -0.1 (-0.1) -0.6 V, -0.8 V, -1 V, -1.5 and -2 V. Frequency range is from 300 MHz to 40 GHz (41 measurement points). To simplify the measurement, D-optimal is only applied on Vds and Vgs which gives 68 experiments and considering frequency the total number of measurements is 2788.

Since the networks with less than 6 hidden neurons are not able to fit to the measured data the minimum number of hidden layer's neurons (N_h) is assumed to be 6. Fig. 1

shows the TSE and penalty terms ($\alpha^4\hat{\sigma}_0^2 k/N$) and Fig. 2 shows MPSE as functions of N_h, and α, respectively. In this verification example α of 0.4 is selected which locates the $MPSE_{min}$ at 14 neurons in the hidden layer.

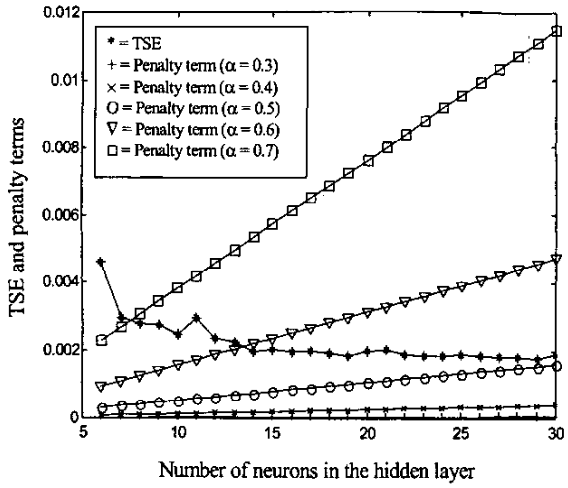

Fig. 1. TSE and penalty terms as function of number of neurons in the hidden layer.

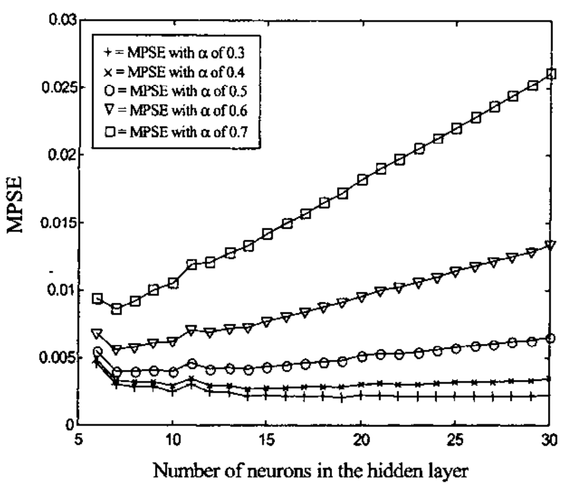

Fig. 2. MPSE as a function of number of neurons in the hidden layer (N_h) and penalty factor (α).In this verification example α of 0.4 is selected which locates the $MPSE_{min}$ at 14 neurons in the hidden layer.

Fig. 3 illustrates predictions of S-parameters by the neural model and measured S-parameters for the bias point of Vds = 4.5 V and Vgs = - 0.2 V which is not used as a training data (the transistor at this bias point is not in the active working region so the magnitude of S11 is smaller than one).

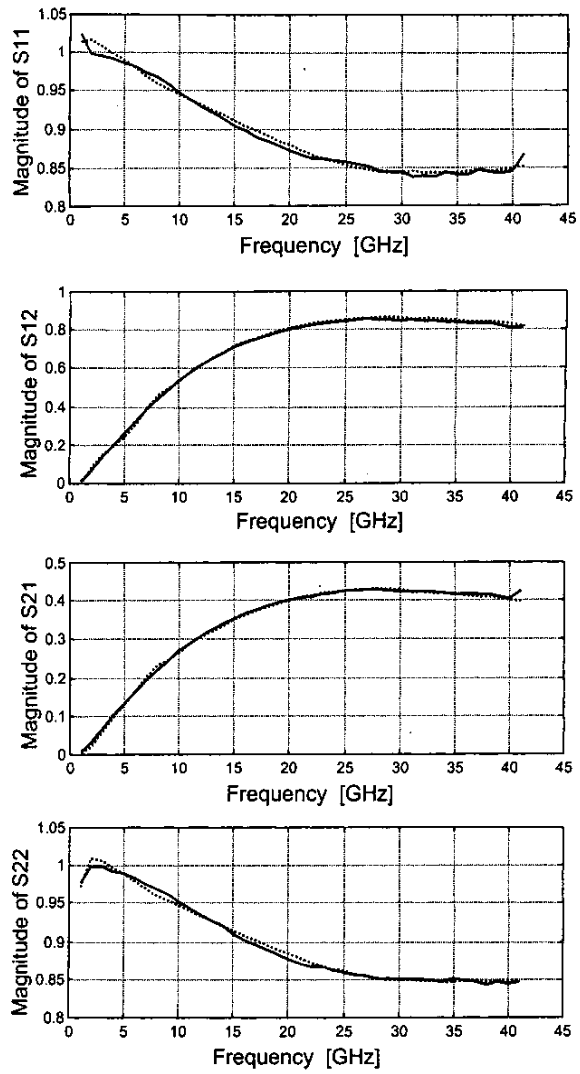

Fig. 3. Predictions of the magnitudes of S-parameters by the neural model compared with measurements for the bias point of Vds = 4.5 V and Vgs = - 0.2 V. Dashed-line for the neural model results and solid-line for measurement results.

To compare the results the average errors for this model and a model trained with the same method but with 24 neurons in the hidden layer are 1.46 % and 1.29 % for the used training data (2788 measurement points) and 2.55 % and 2.48 % for all the available measured data (8200 measurement points), respectively. It means the average error related to the predictions of new data for the neural model with 14 neurons in the hidden layer is smaller.

To compare this model to the standard MLP with training algorithm of error back propagation all the measurements data (8200 measurements) have been used to train a neural network with 24 neurons in the hidden layer. The average error for the standard MLP network is 1.63 % but as the Fig. 5 shows the predictions by this neural network is not smooth and is effected by the noise.

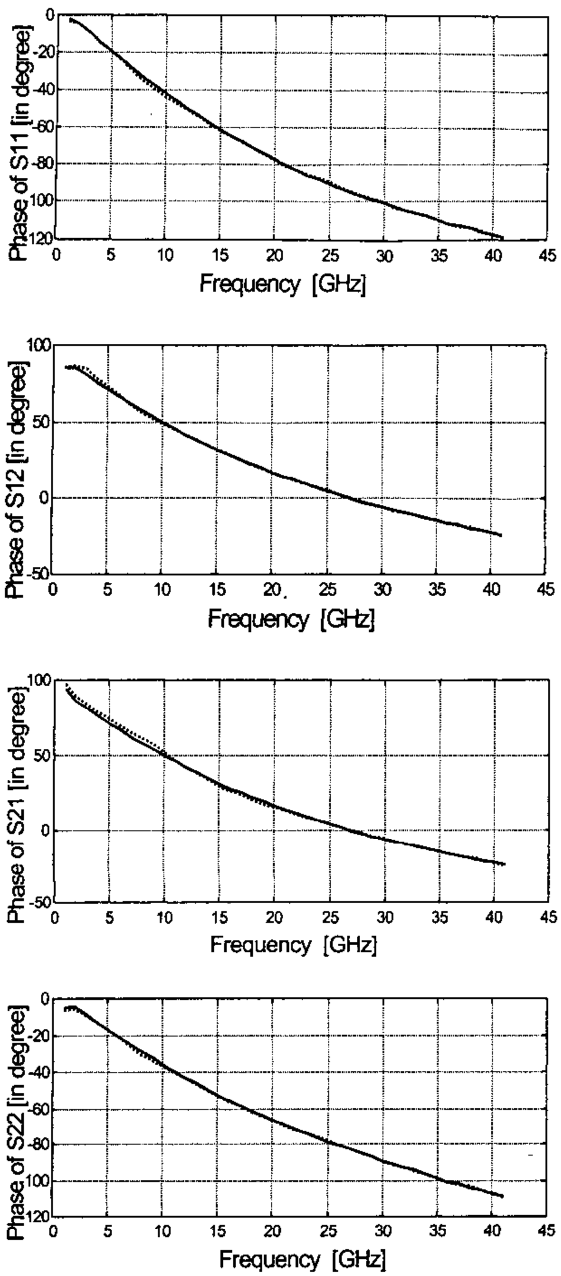

Fig. 4. Prediction of the phase of S-parameters by the neural model compared with measurements for the bias point of Vds = 4.5 V and Vgs = - 0.2. Dashed-line for the neural model results and solid-line for measurement results.

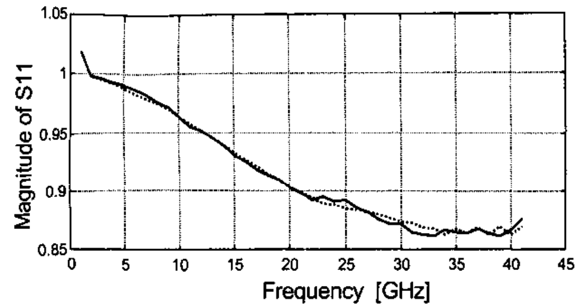

Fig. 5. Prediction of the magnitude of S11 compared with measurement for the same bias point used in Figs. 3 and 4. Dashed-line for the neural model results and solid-line for measurement results.

IV. Conclusion

A systematic method has been introduced to construct a reliable neural model for microwave active devices. Although a small number of training data has been used and the neural network has a simple structure, the model gives very smooth and accurate predictions.

Acknowledgement

The authors wish to thanks Mr. Wilfred Mwema, Dr. Friedbert van Raay and Mr. Jürgen Weide for their helpful co-operation. This work has been done within a project of Material and Components of Microsystems Technology which is supported by German Research Foundation (DFG).

References

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

[2] P. M. Watson and K. C. Gupta, "EM-ANN Models for Microstrip Vias and Interconnects in Multilayer Circuits," *IEEE Trans. Microwave Theory and Tech.*, vol. MTT-44, no. 12, pp. 2495-2503, Dec. 1996.

[3] V. K. Devabhaktuni, C. Xi, and Q. J. Zhang, "A Neural Network Approach to the Modeling of Heterojunction Bipolar Transistors from S-parameter Data," *Proc. European Microwave Conf.*, vol. 1, pp. 306-311, Oct. 1998.

[4] Y. L. Huang, T. F. Edgar, D. M. Himmelblau, and I. Trachtenberg, "Constructing a Reliable Neural Network Model for a Plasma Etching Process Using Limited Experimental Process," *IEEE Trans. Semiconductor Manufacturing,* vol. 7, no. 3, pp. 333-345, August 1994.

[5] R. Hetch-Nielsen, *Neurocomputing,* Reading, MA: Addison Wesley, 1990.