Nonlinear AlGaN/GaN HEMT Model Using Multiple Artificial Neural Networks

P. Barmuta, P. Płoński, K. Czuba
Warsaw University of Technology
Warsaw, Poland
Pawel.Barmuta@gmail.com

G. Avolio, D. Schreurs
KU Leuven
Leuven, Belgium

***Abstract*—In this work, a complete nonlinear-transistor-model extraction-method is described. As a case study, the AlGaN/GaN High Electron Mobility Transistor manufactured on SiC substrate is modeled. The parasitic components model is proposed, and its extraction results are presented. Low- and high-frequency large-signal measurement data are involved in order to produce multiple artificial neural networks. The network topologies of multilayer perceptron networks are established automatically. A complete learning procedure using back propagation algorithm is described. A good agreement between the measurement data and the model has been observed.**

Keywords-artificial neural network; temperature; GaN HEMT; nonlinear model

I. INTRODUCTION

During the past few years, AlGaN/GaN devices have been a subject of intensive investigation. Due to considerable bandgap width they are suitable for high-frequency, high-power applications [1]. Accurate nonlinear models are essential to their design process. A number of models have been developed during past two decades. The common modeling techniques are based on closed-form equations [2]. They provide extrapolating capabilities, however their development is very time consuming. Furthermore, extraction of certain model parameters often entails additional extensive measurements. Other conventional nonlinear modeling methods, i.e., look-up tables [3] and Volterra series [4] typically introduce high computational and/or memory requirements. To eliminate the aforementioned drawbacks, some attempts [5] have been made to model the nonlinearities with artificial neural networks (ANN).

The ANNs provide good approximation capabilities, which can model even strongly nonlinear behaviour. According to the universal approximation theorem [6], there always exists a 3-layer multilayer perceptron (MLP) that approximates any nonlinear, continuous function with a desired accuracy. However, the number of required neurons could be large in some cases. Some efforts have been made to develop an automatic algorithm for optimal ANN-neuron-number determination [7]. Since the ANN does not involve any assumption or physical bonds, the sufficient amount of the training data is required. Even though, it could be large for high-dimensional problems, the model never requires any additional measurements as in case of the closed-form equations and physical models.

In this paper, a complete systematic procedure of nonlinear modelling using ANN is presented. In order to minimize the approximation error, multiple MLP ANNs have been combined into one model. As a case study, the AlGaN/GaN High Electron Mobility Transistor working as an A-class amplifier has been modelled. The load-pull measurements of the DUT were performed using two large signal network analysers: one for low frequency (20 kHz – 25 MHz) and the other one for high frequency (4 GHz) range. Voltages and currents of DC and 10 first harmonics were measured (real and imaginary parts) both for the gate and the drain side. In order to ensure the data consistency, the same bias point and signal swing was applied in both frequency ranges. Additionally, the scattering matrix of the transistor was measured in frequency range of 300 MHz to 40 GHz using vectorial network analyser. All the measurements were performed on-wafer on a temperature-controlled chuck. More details on the measurement procedure are given in [8].

The paper is organized as follows: In Section 2 the multilayer perceptron networks are briefly described. In Section 3 the modelling procedure is presented. In Section 4 the modelling results are shown and discussed. Finally the conclusions are stated in Section 5.

II. MULTILAYER PERCEPTRON NETWORK

A model of a single neuron is presented in Fig. 1. Output of the single neuron is computed according to:

$$\mu_i = f\left(\sum_{k=0}^{k=j} w_{ik} x_k\right), \tag{1}$$

where f is called the activation function and w_{ik} are the weights which multiply the neuron inputs. Neurons are connected into layers. A structure of MLP with 2 hidden layers is shown in Fig. 2. The output values of the MLP are computed by feed-forward propagating an input signal from the input neurons, through the hidden layers, to the output neurons [9]. The more layers are involved the less neurons are needed to achieve the desired approximation error. On the other hand, more layers mean more connections, thereby making the learning process more resource consuming.

978-1-61284-1438-2/12/$26.00

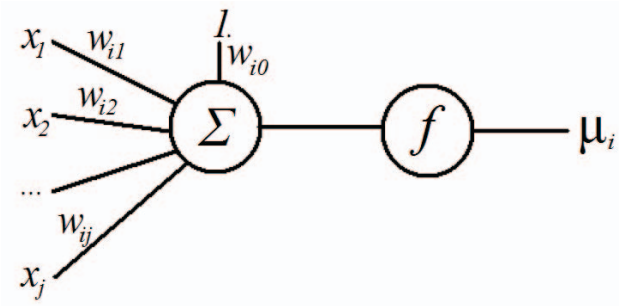


Figure 1. Model of a single neuron

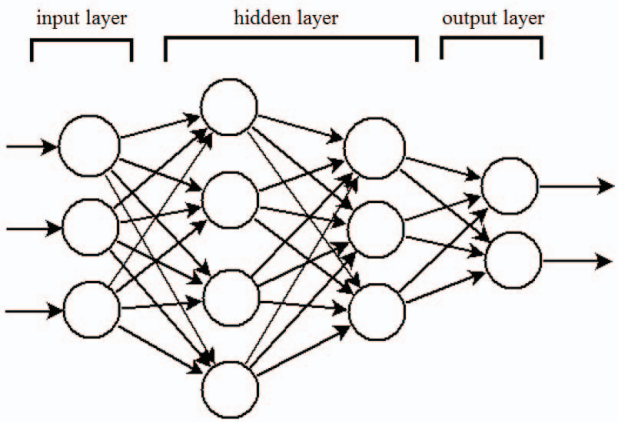


Figure 2. Structure of a multilayer perceptron of type 3-4-3-2 (3 neurons in the input layer, 4 in the first hidden layer, 3 in the second hidden layer and 2 in the output layer).

III. Modelling Procedure

A general active-device model is shown in Fig. 3. It consists of the intrinsic nonlinear active part surrounded by linear parasitic components. The modeling procedure can be described as follows:

A. Parasitics De-embedding

The linear parasitic components model is determined based on the scattering matrix measurements. Since the large signal measurements are performed at significantly lower frequencies, the parasitics model has to provide extrapolation capabilities. Thus, the ANN cannot be used. The parasitics are modeled with the equivalent circuit model shown in Fig. 4.

The model extraction method is derived from the method described in [10]. However, it is extended by introducing substrate related parasitics C_{gg}, R_{gg}, C_{dd}, R_{dd}. Additional parameters makes the fitting target function nonlinear, thereby the linear regression cannot longer be used. For each equation describing the DUT a nonlinear regression model is constructed. They are fitted simultaneously with the biggest trust region method. The extraction results are presented in Table I. All the values are extracted from the data measured at the pinch-off "cold" condition (gate-to-source voltage value V_{gs} = -6 V, drain-to-source voltage value V_{ds} = 0 V). All the obtained values are positive and have physical interpretation.

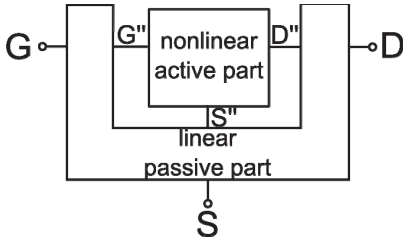


Figure 3. General active device model.

B. Current Generators Modelling

The nonlinear model has to be extracted from the large-signal measurements data. After de-embedding the parasitic components, the reference plane is shifted to the intrinsic nonlinear part. A black-box approach can be used, since for the large-signal measurement data no extrapolation is required. However, it cannot be modeled within one ANN, since the data do not show straight-forward dependency on the stimuli. Therefore, in order to achieve better fitting quality, the intrinsic nonlinearities are described with the model shown in Fig. 5. It consists of two nonlinear current generators responsible for the drain current (I_d) and the gate current (I_g), and 2 nonlinear charge sources responsible for the capacitive effects modeling. Both, the current generators and the charge sources are modeled with the MLP ANNs.

To minimize both the neuron and the connection number, the 4-layer MLP are used for each ANN. In the hidden layers the sigmoidal activation function was used, in the output layer activation function was linear. The topology of each network was established automatically according to [7]. The conjugated scaled gradient algorithm was used for the network training. The behavior of the current generators depends on the temperature and traps state in the device, as well as on the bias point. Trapping phenomena are frequency dependent, since they have different time constants. However, traps gradually stop to follow the signal swing, as the excitation frequency

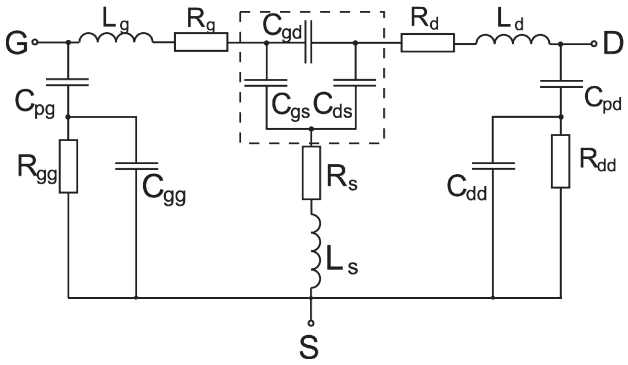


Figure 4. AlGaN/GaN high electron mobility transistor parasitics equivalent circuit model extracted at the pinch-off "cold" conditions

TABLE I. PARASITIC COMPONENTS VALUES EXTRACTED AT THE PINCH-OFF "COLD" CONDITIONS.

Element	Value
C_{gd}	234 fF
C_{gs}	156 fF
C_{pg}	61 fF
R_{gg}	317 Ω
C_{gg}	70 fF
C_{ds}	187 fF
C_{pd}	65 fF
R_{dd}	338 Ω
C_{dd}	114 fF
R_d	0.69 Ω
R_s	67 mΩ
R_g	3.35 Ω
L_d	119 pH
L_s	3.47 pH
L_g	143 pH

grows. Meanwhile, at low frequencies the capacitances (and the charge sources alike) are not pronounced. Therefore, the current generators should be extracted from the low-frequency large-signal data, for which the traps state can be perceived as constant.

For the low-frequency measurements one can directly control the DC values and the first harmonic of the voltages. Therefore, those quantities were chosen as the ANN input. The data was normalized to zero to one range prior to modeling, in order to improve the approximation quality. The low-frequency and high-frequency data were combined for the normalization process, by subtracting the low-frequency ANN output from the high-frequency data. Moreover, all the data were shifted in phase so as to obtain 0° phase for the first harmonic of the V_{ds}. Without the phase normalization the data would not be eligible for comparison, and the learning process would be disturbed. The network can be trained either in the time domain data or in the frequency domain. Since the frequency domain requires less data, the spectrum approach has been chosen for modeling purposes. Since the I_g values are very small, on the brink of the measurement system capabilities, the current generator for the I_g was not modeled. It would result in an erroneous simulation for the high frequency data.

C. Charge Sources

Unlike the current generators, the charge sources are frequency dependent. The impact of nonlinear capacitances is especially pronounced at high frequencies. Therefore, the large-signal high-frequency data should be used to model them. Prior to charge sources modeling, the current sources have to be de-embedded. Thus, the response of the previously trained current source ANN is determined for the high-frequency data. Subsequently, the obtained current is subtracted from the high-frequency current. The results again have to be normalized to zero to one range.

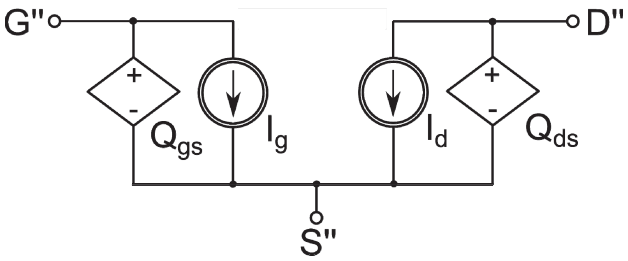


Figure 5. Equivalent circuit for the nonlinear, intrinsic elements.

As opposed to the low-frequency measurements, at high frequencies one has no direct control of the voltage and phase. What is more, after the de-embedding of the current generators, the charge sources depend on all of the harmonics. As a result, the ANN has to be trained using DC and all the harmonics of the drain and gate voltages as the input. Drain- and gate-side charge source networks could be fitted within single network. However, it would result in a large weight matrix and consequently in a very resource consuming learning process. Although, the weight matrix can be reduced by neuron connection removal algorithms [12], it would still put high requirements on computational time. Data which shows similar behavior should be fitted together in one network. Therefore, in the first place the charge source ANNs for drain and gate sides were established individually. Also, the gate-to-source charge source (Q_{gs}) ANN was also split into two separate networks, in order to achieve better approximation quality within the smaller effort. What is more, the I_g data shows strong nonlinearity, which manifests itself with large even harmonics amplitudes.

IV. RESULTS AND DISCUSSION

The I_d current generator ANNs topology is 9-9-7-7. The drain side charge source ANN topology is 41-13-11-21. The gate side charge source ANN topologies are 41-9-7-11 and 41-9-7-10 for the even and odd harmonics respectively.

550 samples of the low-frequency data and 200 samples of high-frequency data are used for the ANN training. 25 samples of high-frequency data are used for verification. The learning process error (E) of all the networks is shown in Fig. 6. The error is normalized to the sample number according to the:

$$E = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2, \qquad (1)$$

where y is the ANN output, t is the desired training value and N is the number of samples. All the training processes show convergence to the minimum error value. It can be seen that the

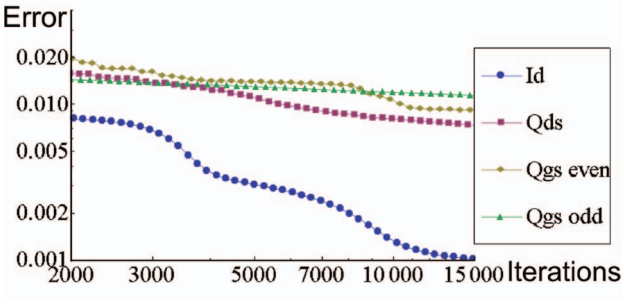


Figure 6. The normalized error of the artificial neural networks as a function of training algorithm iteration.

smallest errors occurs for the drain-side data learning. This is caused by the larger signal-to-noise ratio in the I_d measurements. The I_d current generator ANN shows significantly smaller training error than the charge sources ANNs. It may result from a better control on the instantaneous voltage values in the lower frequencies range. The instantaneous voltages values vary more at high-frequency measurements, which causes the data to be less consistent.

The cross-correlation between the I_d current source ANN output and the desired training values are shown in Table II. ANN output data correspond very well to the one used for learning, as all the cross-correlation coefficients are above 0.98. The cross-correlation between the charge sources ANNs outputs and the desired training values are shown in Table III. Cross-correlation values are bigger for the drain-to-source charge source (Q_{ds}) ANN, which is consistent with Fig. 6. The odd harmonics approximation is worse for both of the charge source networks. As the device behavior is strongly nonlinear, it manifests itself in high even harmonic amplitudes. Therefore, despite the normalization, the even harmonics are approximated better than the odd ones.

A comparison between the verification data and the model is shown in Fig. 7. A good correspondence between the simulation and measurements can be observed. One must bear in mind that the data was approximated in the frequency and not in the time domain.

V. CONCLUSIONS

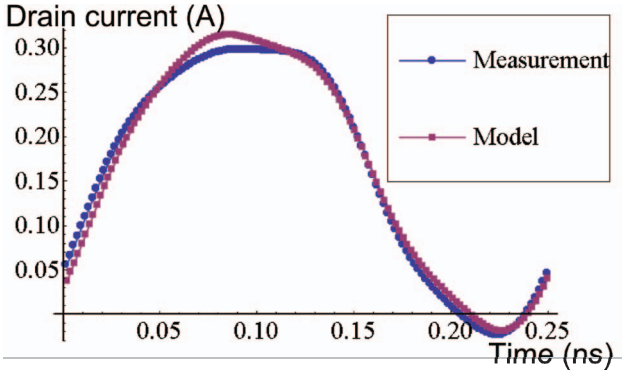
A complete procedure of nonlinear model extraction from the large-signal measurements data has been reported. A particular emphasis has been placed on the modeling of the intrinsic nonlinearities with multiple artificial neural networks. The 4-layer perceptron network with algorithm for automatic adjustment of the neuron number was used. The learning procedure was implemented with back propagation algorithm called scaled conjugated gradient. Network extraction and approximation improving techniques have been described. The proposed procedure has been applied to establish the nonlinear model for AlGaN/GaN high electron mobility transistor. A good correspondence between the measurement data and the simulation has been observed.

TABLE II. CROSS-CORRELATION BETWEEN THE ID ANN OUTPUT AND THE TRAINING DATA.

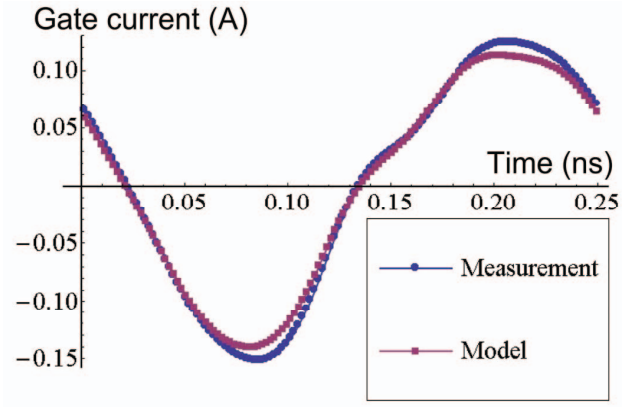
DC	1st harmonic		2nd harmonic		3rd harmonic	
	Re	*Im*	*Re*	*Im*	*Re*	*Im*
0.9978	0.9955	0.9834	0.9971	0.9822	0.9938	0.9851

TABLE III. CROSS-CORRELATION BETWEEN THE ID ANN OUTPUT AND THE TRAINING DATA.

Harmonic no.	$Re(Q_{ds})$	$Im(Q_{ds})$	$Re(Q_{gs})$	$Im(Q_{gs})$
DC	0.9538		0.3740	
1	0.9812	0.9526	0.3864	0.5161
2	0.9873	0.9612	0.9869	0.9864
3	0.6405	0.9753	0.5015	0.6511
4	0.8884	0.9229	0.9514	0.9394
5	0.6465	0.7023	0.7653	0.2684
6	0.9889	0.9935	0.7766	0.4554
7	0.5460	0.5086	0.7949	0.7941
8	0.9703	0.9823	0.7284	0.6639
9	0.8377	0.7945	0.8557	0.7599
10	0.6242	0.7379	0.9744	0.9706



(a)



(b)

Figure 7. Comparison between data and model; (a) drain, (b) gate current. Excitation frequency 4 GHz.

ACKNOWLEDGMENT

The authors would like to thank FWO-Vlaanderen and KU Leuven GOA for funding the project. P.B would like to thank Dr. Wojciech Wiatr and Dr. Arkadiusz Lewandowski for valuable suggestions.

REFERENCES

[1] J.C. Zolper, "Wide bandgap semiconductor microwave technologies: from promise to practice," in IEDM Tech. Dig., pp. 389–392, Dec. 1999.

[2] I.Angelov, L.Bengtsson, and M.Garcia, "Extensions of the chalmers nonlinear HEMT and MESFET model," IEEE Transactions on Microwave Theory and Techniques, vol. 44, pp. 1664 – 1674, Oct. 1996.

[3] D. Root, S. Fan, and J. Meyer, "Technology independent large-signal nonquasi-static FET models by direct construction from automatically characterized device data," in Proc. 21st Europ. Microwave Conf., Stuttgart, Germany, vol. 2, pp. 923–927, Sept. 1991.

[4] S.A. Maas and A. Crosmun, “Modeling the gate I V characteristic of a GaAs MESFET for Volterra-series analysis,” IEEE Trans. Microwave Theory Tech., vol.37, pp. 1134–1136, July 1989.

[5] G. Orengo, “Advanced neural network techniques for GaN-HEMT dynamic behavior characterization,” European Microwave Integrated Circuits Conference, pp. 249 – 252, Sept. 2006.

[6] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators," Neural Networks, vol. 2, pp. 359-366, 1989.

[7] R. Sulej, K. Zaremba, and K. Kurek, “Dynamic topology adjustment algorithm for MLP networks," Articial Intelligence and Soft Computing, 2006.

[8] P. Barmuta, et al.,”Temperature dependent vector large-signal measurements,” Integrated Nonlinear Microwave and Millimetre-Wave Circuits Workshop, pp. 1-4, Apr. 2011

[9] C.M. Bishop, “Neural Networks for Pattern Recognition”, Clarendon Press, Oxford, 1996.

[10] G. Crupi; D.M.M.-P. Schreurs, A. Raffo, A. Caddemi, G. Vannini, “A New Millimeter-Wave Small-Signal Modeling Approach for pHEMTs Accounting for the Output Conductance Time Delay,” IEEE Trans. on Microwave Theory and Techniques, vol. 56, pp 741 – 746, Apr. 2008.

[11] A. Raffo, et al. “Nonlinear dispersive modeling of electron devices oriented to GaN power amplifier design,” IEEE Trans. Microw. Theory and Tech., Vol. 58, no. 4, pp. 710-718, Apr. 2010.

[12] B. Hassibi and D.G. Stork and Stork, "Second Order Derivatives for Network Pruning: Optimal Brain Surgeon", Advances in Neural Information Processing Systems, pp. 164-171, 1993.