Neural Networks for Large- and Small-Signal Modeling of MESFET/HEMT Transistors

Marcelino Lázaro, *Student Member, IEEE*, Ignacio Santamaría, *Member, IEEE*, and Carlos Pantaleón, *Member, IEEE*

***Abstract*—In this paper, we present a comparative study of three neural networks-based solutions for large- and small-signal modeling of MESFET and HEMT transistors. The first two neural architectures are specific for this modeling problem: the generalized radial basis function (GRBF) network, and the smoothed piecewise linear (SPWL) model. These models are compared with the well-known multilayer perceptron (MLP) network. Results are presented for both the large- and small-signal regimes separately. Finally, a global model is proposed that is able to accurately characterize the whole behavior of the transistors. This model is based on a simple combination of the best models obtained for the two kinds of regimes.**

***Index Terms*—Intermodulation, large-signal, microwave transistors, neural networks, nonlinear modeling, small-signal.**

I. Introduction

THE DESIGN of microwave and millimeter-wave circuits and the increasing integration of hybrid and monolithic circuits have reinforced the need of accurate large- and small-signal device models to improve the performance of these circuits and to minimize the number of design and fabrication steps required. For the small-signal models, to reproduce the third-order intermodulation behavior is quite a difficult and common task (amplifiers working below the 1 dB compression point and mixers excited by small RF signals when compared to the local oscillator are typical examples). In this case, it is necessary to approximate not only the current–voltage ($I-V$) characteristic, but also its derivatives up to the third order [1]. For the large-signal models it is possible to characterize the RF large-signal behavior approximating the pulsed dc behavior of the devices [2].

Conventional nonlinear techniques applied to device modeling, such as closed-form equations [3], [4], Volterra series [5], or the use of look-up tables [6], present high memory requirements or a high computational burden. Moreover, they are not conceived to model the intermodulation behavior.

Recently, some attempts have been made to model the nonlinear behavior of active devices and circuits by using neural networks [7], [8]. Neural networks have the capability of approximating any nonlinear function and the ability to learn from experimental data; therefore, they are good candidates to solve device-modeling problems. However, practically all these neural approaches only consider the use of the multilayer perceptron (MLP) and, in this case, the memory requirements to give a good approximation, and the computational requirements to carry out the training of the network are also high.

In order to avoid these problems we have recently proposed two models: the generalized radial basis functions (GRBF) network [9], for small-signal modeling with the capability of reproducing the $I-V$ derivatives, and the smoothed piecewise linear (SPWL) model [10] for large-signal regimes. Both models require a low number of parameters, and their computational requirements to train the models are lower than those of the above-mentioned methods. In this paper, we present a comparative study of these new architectures (as well as the MLP), applied to small- and large-signal device modeling. The conclusions of this study are used to propose a global model formed by merging in an appropriate way the best small- and large-signal models.

The paper is organized as follows. In Section II the problem of modeling microwave transistors is stated. In Section III the GRBF and SPWL models are described. In Section IV the main results are presented, and in Section V a global model is proposed, to characterize the whole device behavior. Finally, in Section VI, the main conclusions are reported.

II. Modeling of MESFET and HEMT Transistors

In this section we state the problems encountered when modeling microwave devices such as MESFET or HEMT transistors. Fig. 1 shows the most widely accepted equivalent nonlinear circuit of the FET transistor in its saturated region. Here, we concentrate on the modeling of the drain-to-source current $I_{\rm ds}$ static nonlineality, but it is simple to consider the dynamic elements of the equivalent circuit. Generally, to model a transistor, there are two clearly different kinds of regimes, the large- and the small-signal regimes.

A. Small-Signal Modeling of Transistors

In a MESFET or HEMT, the predominant nonlinear element is the drain-to-source current $I_{\rm ds}$, which depends on the drain-to-source $V_{\rm ds}$, and the gate-to-source $V_{\rm gs}$, bias voltages. This dependence is denoted as the $I-V$ characteristic. As it is shown in [1], the nth-order intermodulation output power varies fundamentally as the square of the nth derivative of the $I-V$ characteristic with respect to the drain-to-source and gate-to-source voltages. Therefore, if we want to be able to model the small-signal intermodulation behavior, our model must accurately fit not only the nonlinear function but also its derivatives. In particular, when we apply a small-signal RF input around a bias point, the drain current $I_{\rm ds}$ depends on the bias point ($V_{\rm ds}$, $V_{\rm gs}$) and on the instantaneous small-signal

Manuscript received May 4, 2000; revisedAugust 20, 2001.

The authors are with DICOM, ETSII y Telecom, University of Cantabria, 39005 Santander, Spain.

Publisher Item Identifier S 0018-9456(01)10938-1.

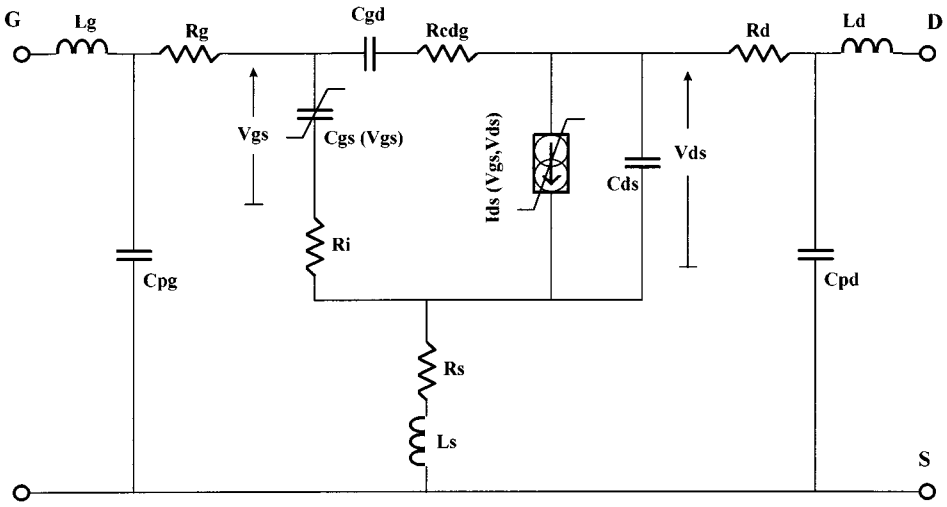

Fig. 1. Nonlinear equivalent circuit of a MESFET transistor.

voltages ($v_{\rm ds}$, $v_{\rm gs}$). We can approximate $I_{\rm ds}$ by the following truncated Taylor series expansion:

$$I_{\rm ds} = I_{\rm dso} + G_{\rm m} v_{\rm gs} + G_{\rm ds} v_{\rm ds} + G_{\rm m2} v_{\rm gs}^2 + G_{\rm md} v_{\rm ds} v_{\rm gs} + G_{\rm d2} v_{\rm ds}^2 + G_{\rm m3} v_{\rm gs}^3 + G_{\rm m2d} v_{\rm ds} v_{\rm gs}^2 + G_{\rm md2} v_{\rm ds}^2 v_{\rm gs} + G_{\rm d3} v_{\rm ds}^3 \quad (1)$$

where $I_{\rm dso}$ is the dc drain current and ($G_{\rm m}$, ..., and $G_{\rm d3}$) are coefficients related to the nth-order derivatives of the I–V characteristic evaluated at the bias point. For instance, the $G_{\rm md2}$ coefficient is defined as

$$G_{\rm md2} = \frac{1}{2} \frac{\partial^3 I_{\rm ds}}{\partial v_{\rm gs} v_{\rm ds}^2}\bigg|_{(V_{gs}, V_{ds})} \quad \text{in } v_{\rm ds} = v_{\rm ds} = 0. \quad (2)$$

Therefore, our small-signal modeling problem consists in fitting a function (model) $\boldsymbol{G}_{SS}$: $\Re^2 \rightarrow \Re^{10}$, which approximates the nonlinear mapping from the input space of bias voltages $\boldsymbol{V} = (V_{\rm ds}, V_{\rm gs})$, to the output space of model parameters $\boldsymbol{G}_{SS} = (I_{\rm dso}, G_{\rm m}, G_{\rm ds}, G_{\rm m2}, G_{\rm md}, G_{\rm d2}, G_{\rm m3}, G_{\rm m2d}, G_{\rm md2}, G_{\rm d3})$. Once this model is available, $I_{\rm ds}$ is reconstructed by (1).

B. Large-Signal Modeling of Transistors

The large-signal behavior of MESFET or HEMT transistors is governed by the nonlinear dynamic pulsed I–V characteristic that depends on the quiescent bias point [2]. An accurate representation of this dc characteristic of pulsed measurements, dependent on the bias point, can provide a suitable representation of the RF large-signal behavior of the transistor [2]. Therefore, in this case, the drain current $I_{\rm ds}$ depends on the bias point $(V_{\rm ds}, V_{\rm gs})$ and on the pulsed voltages $(v_{\rm ds}, v_{\rm gs})$ applied over the bias point. Note that here we are using pulsed dc samples to obtain a suitable representation of the large-signal RF behavior. Now, our large-signal modeling problem consists in obtaining a function $\boldsymbol{G}_{LS}$: $\Re^4 \rightarrow \Re^1$, which approximates the nonlinear mapping from the input space of bias and pulsed voltages $\boldsymbol{V} = (V_{\rm ds}, V_{\rm gs}, v_{\rm ds}, v_{\rm gs})$ to the output space $\boldsymbol{G}_{LS}(\boldsymbol{V}) = I_{\rm ds}$.

III. Proposed Networks

In this section, we describe the GRBF and SPWL networks that we have proposed to solve the above modeling problems.

A. Generalized Radial Basis Function (GRBF) Network

The GRBF network is an extension of the radial basis function (RBF) network that relaxes the radial constraint for the basis kernels allowing different variances for each dimension of the input space [9]. In this way, it is possible to reduce the number of required basis functions, and therefore the number of parameters. To perform a general $\boldsymbol{G}$: $\Re^J \rightarrow \Re^M$ mapping, the kth output of the GRBF network is given by

$$G_k(\boldsymbol{V}) = \sum_{i=1}^{I} g_i(\boldsymbol{V}) \quad (3)$$

where i indexes the different GRBF units $g_i(V) = \lambda_{ik} o_i(V)$ and $o_i(V)$ is the activation function of each unit

$$o_i(\boldsymbol{V}) = \prod_{j=1}^{J} \exp - \frac{(V_j - \mu_{ij})^2}{2\sigma_{ij}^2} \quad (4)$$

where V_j is the jth element of input vector V.

The network is initialized by a variant of the orthogonal least squares (OLS) algorithm [11], which is able to work with elliptical kernels. The error function E to be minimized is the quadratic error. The variances and the centers of the network are adapted by using the following equations of the gradient

$$\frac{\partial E}{\partial \sigma_{ij}} = -2 \sum_p \sum_k \cdot e_k(\boldsymbol{V}_p) o_i(\boldsymbol{V}_p) \lambda_{ik} \frac{1}{\sigma_{ij}} \left(\frac{V_{pj} - \mu_{ij}}{\sigma_{ij}} \right)^2 \quad (5)$$

$$\frac{\partial E}{\partial \mu_{ij}} = -2 \sum_p \sum_k \cdot e_k(\boldsymbol{V}_p) o_i(\boldsymbol{V}_p) \lambda_{ik} \frac{1}{\sigma_{ij}} \left(\frac{V_{pj} - \mu_{ij}}{\sigma_{ij}} \right) \quad (6)$$

where p indexes the input patterns, k the output dimensions, and $y_k(\boldsymbol{V}_p)$ and $e_k(\boldsymbol{V}_p)$ are the desired output and the network error, respectively, of the kth output dimension for the pth input pattern. With the centers and variances fixed, the dependence with the λ_{ik} parameters is linear, and their optimum values are

easily calculated by least squares. This adaptation process is iterated until a suitable error is reached.

B. The Smoothed Piecewise Linear (SPWL) Model

The SPWL model is an extension of the well-known canonical piecewise linear model proposed by Chua [12]. This model, as it is shown in [13], can be seen as a neural network. Basically, the model implements a general mapping $\boldsymbol{G}$: $\Re^M \rightarrow \Re^N$ as follows:

$$\boldsymbol{G}(\boldsymbol{V}) = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{V} + \sum_{i=1}^{\theta} \boldsymbol{c_i} \left| \langle \boldsymbol{\alpha_i}, \boldsymbol{V} \rangle - \beta_i \right| \tag{7}$$

where

$\boldsymbol{V}$ and $\boldsymbol{\alpha_i}$	vectors of dimension M;
$\boldsymbol{a}$ and $\boldsymbol{c_i}$	vectors of dimension N;
$\boldsymbol{B}$	$N \times M$ matrix;
β_i	scalar;
$\langle , \rangle$	inner product.

The model divides the input space into different regions by means of several boundaries implemented by hyperplanes of dimension $M-1$ (defined by the expression inside the absolute value function), and it carries out the function approximation by means of the combination of hinging hyperplanes of dimension M, which are the result of joining linear hyperplanes over the boundaries defined in the input space. In any region the model is composed by a linear combination of linear hyperplanes, and the transitions at the boundaries are governed by the absolute value function. Therefore, this model inherits some properties from the absolute value function: it is continuous but not derivable along the boundaries. Moreover, the second- and higher order derivatives are zero except at the boundaries where they are discontinuous.

To overcome this drawback, the SPWL model substitutes the absolute value function for a smooth and derivable function in order to smooth the transition at the boundaries dividing the input space [10]. Several possibilities exist to smooth the absolute value function allowing, at the same time, a parametric control of the "sharpness" of the transition. We have proposed the following smoothing function:

$$lch(x, \gamma) = \frac{1}{\gamma} \ln(\cosh(\gamma x)) \tag{8}$$

where γ is a parameter that allows controlling the smoothness of the transition. Finally, the proposed SPWL model is given by

$$\boldsymbol{G}(\boldsymbol{V}) = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{V} + \sum_{i=1}^{\theta} \boldsymbol{c_i} lch\left(\langle \boldsymbol{\alpha}_i, \boldsymbol{V} \rangle - \beta_i, \gamma_i\right). \tag{9}$$

The training of this network is carried out by means of an optimization method equivalent to the method proposed by Chua for the canonical piecewise linear model [12].

Let us consider that we want to approximate a mapping $\Re^M \rightarrow \Re$ using a set of N input–output samples $(\boldsymbol{V}_l, y_l)$, $l = 1, \ldots, N$ with $\boldsymbol{V}_l = (V_{1,l}, V_{2,l}, \ldots, V_{M,l})$. Assuming that $\alpha_{t,M} \neq 0$, we can eliminate one coefficient from each boundary by rewriting $\langle \boldsymbol{\alpha_i}, V \rangle - \beta_i$, as

$$b_i(\boldsymbol{V}) = m_{i,1}V_1 + m_{i,2}V_2 + \cdots + m_{i,M-1}V_{M-1} - V_M + t_i \tag{10}$$

where $b_i(\boldsymbol{V})$ denotes the ith boundary evaluated at $\boldsymbol{V}$. Finally, taking into account that $\boldsymbol{B}$ and $\boldsymbol{a}$ in (9) are now a vector $\boldsymbol{b} = (b_1, \ldots, b_M)^T$, and a scalar a, respectively, our generic SPWL model, with θ boundaries, can be written as

$$G(\boldsymbol{V}) = a + \mathbf{b}^T\boldsymbol{V} + \sum_{i=1}^{\theta} c_i\, lch(b_i(\boldsymbol{V}), \gamma). \tag{11}$$

The model parameters can be grouped into two vectors: $\boldsymbol{z_p}$ grouping the coefficients associated to the linear combination of components

$$z_p = (a, b_1, \ldots, b_M, c_1, \ldots c_\theta)^T \tag{12}$$

and $\boldsymbol{z_r}$, grouping the parameters defining the boundaries of the domain space

$$\boldsymbol{z_r} = (m_{1,1}, \ldots, m_{1,M-1}, \ldots, m_{\theta,1}, \ldots m_{\theta,M-1}, t_1, \ldots t_\theta)^T. \tag{13}$$

The error function to be minimized is given by

$$E(\mathbf{z}_p, z_r) = \sum_{l=1}^{N} \cdot \left(y_l - \left(a + \boldsymbol{b}^T\boldsymbol{V_l} + \sum_{i=1}^{\theta} c_i\, lch(b_i(\boldsymbol{V_l}), \gamma) \right) \right)^2. \tag{14}$$

The algorithm begins by fixing the initial location of each partition boundary, i.e., the vector $\boldsymbol{z_r}$. Generally, they are chosen randomly. Then, the approximation error $E(\boldsymbol{z_p}, \boldsymbol{z_r})$ is a quadratic function of $\boldsymbol{z_p}$, and its minimum is easily obtained by least squares.

Once the optimal $\boldsymbol{z_p}$ parameters (for a given initial $\boldsymbol{z_r}$ partition) are calculated, the algorithm estimates a new optimal partition $\boldsymbol{z_r}$. This partition is found by calculating the gradient $\boldsymbol{g}$ and the Hessian $\boldsymbol{Y}$, which specify the optimal searching direction to modify $\boldsymbol{z_r}$ according to

$$\boldsymbol{s} = -\boldsymbol{Y}^l\boldsymbol{g}. \tag{15}$$

The gradient $\boldsymbol{g}$ and the Hessian $\boldsymbol{Y}$ are given by

$$\boldsymbol{g} = 2\boldsymbol{K}\boldsymbol{G}\boldsymbol{e}, \tag{16}$$

$$\boldsymbol{Y} = 2\boldsymbol{K}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{K} + 2\boldsymbol{K}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{z}_r}\boldsymbol{e} \tag{17}$$

where $\boldsymbol{e} = (e_1, \ldots, e_N)^T$ is the vector of errors. $\boldsymbol{K}$ is given by

$$\boldsymbol{K} = diag\Bigg(\underbrace{c_1, \ldots, c_1}_{M-1\ terms} \underbrace{c_2, \ldots, c_2}_{M-1\ terms} \cdots \underbrace{c_\theta, \ldots, c_\theta}_{M-1\ terms}, \; c_1, c_2, \ldots, c_\theta \Bigg) \tag{18}$$

TABLE I
SMALL-SIGNAL RESULTS USING INDIVIDUAL NETWORKS FOR EACH OUTPUT. THE NUMBER BETWEEN PARENTHESES INDICATES THE NUMBER OF BASIS FUNCTIONS EMPLOYED FOR EACH MODEL

	N_{param}	I_{ds}	G_{d2}	G_{d3}	G_{ds}	G_{m1}	G_{m2}	G_{m2d}	G_{m3}	G_{md}	G_{md2}
MLP(6)	25x10	38.4	24.9	20.1	37.6	42.4	38.4	21.5	21.2	32.4	20.1
GRBF(5)	25x10	44.3	29.1	21.5	38.8	43.7	34.3	22.2	25.2	34.3	22.2
SPWL(6)	22x10	45.4	28.8	19.8	43.3	46.5	36.7	22.7	23.0	35.2	19.9

TABLE II
SMALL-SIGNAL RESULTS USING A SINGLE NETWORK. THE NUMBER BETWEEN PARENTHESES INDICATES THE NUMBER OF BASIS FUNCTIONS EMPLOYED FOR EACH MODEL

	N_{par}	I_{ds}	G_{d2}	G_{d3}	G_{ds}	G_{m1}	G_{m2}	G_{m2d}	G_{m3}	G_{md}	G_{md2}
MLP(8)	114	29.8	17.9	17.7	34.0	36.2	29.0	20.2	19.6	24.6	17.6
GRBF(8)	114	29.7	17.5	18.5	30.4	31.0	26.2	22.4	22.0	24.2	18.9
SPWL(7)	115	36.1	17.6	17.6	33.6	39.6	27.1	20.4	17.8	29.1	18.5
MLP(11)	153	30.9	22.8	19.2	34.6	37.9	29.0	23.2	21.4	25.8	18.4
GRBF(11)	154	32.2	19.4	19.7	33.7	36.0	31.2	23.5	23.0	29.1	21.5
SPWL(10)	151	42.3	27.6	18.8	41.2	41.6	31.7	22.3	21.0	32.1	22.0

and $\boldsymbol{G}$ is the following matrix:

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}^1 \\ \vdots \\ \boldsymbol{G}^\theta \\ \boldsymbol{P} \end{bmatrix} \tag{19}$$

where $\boldsymbol{G}^{\mathrm{k}}$ are $M - 1 \times N$ matrices with elements $g_{i,j}^k = V_{i,j} \tanh(\gamma b_k(\boldsymbol{V}_j))$, and $\boldsymbol{P}$ is the $\theta \times N$ matrix with elements $p_{i,j} = \tanh(\gamma b_i(\boldsymbol{V}_j))$. The second term of (17) involves the second derivative of the SPWL model $\operatorname{sech}^2(b_i(\boldsymbol{V}_l))$, which is a localized function along the boundaries: only points close to the boundaries contribute to this term. In practice, it has been observed that a great computational saving (without any noticeable degradation) can be achieved by dropping out this term, that is, we use $\boldsymbol{Y} = 2\boldsymbol{K}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{K}$. Once the search direction (15) has been calculated, the new boundaries are estimated as

$$\boldsymbol{z}_r = \boldsymbol{z}_r + \alpha \boldsymbol{s} \tag{20}$$

where $\alpha = \arg\min(E(\boldsymbol{z_p}, \boldsymbol{z_r} + \alpha s))$. With this new partition, the process is repeated: the optimal coefficients $\boldsymbol{z_p}$ are calculated for these new boundaries, and then the optimal partition is re-estimated again, until a given error is reached.

IV. RESULTS

In this section, we present the results obtained to model the NE72084 MESFET from experimental measurements with the two networks described in the previous section, and they are compared with those provided by the MLP network as an appropriate reference. Moreover, a brief study of the computational burden needed to train the models is also presented. Although here we are only presenting results for a MESFET transistor, the proposed models can also be applied to HEMT transistors.

A. Small-Signal Results

For the small-signal model, we dispose of a set of measurements of the parameters included in the Taylor series (1) $(I_{\mathrm{dso}}, G_{\mathrm{m}}, G_{\mathrm{ds}}, G_{\mathrm{m2}}, G_{\mathrm{md}}, G_{\mathrm{d2}}, G_{\mathrm{m3}}, G_{\mathrm{m2d}}, G_{\mathrm{md2}}, G_{\mathrm{d3}})$, which compose the output of the small-signal model. These parameters were measured at different bias voltages in the following grid: V_{ds} from 3 to 6 V in steps of 0.25 V, and V_{gs} from -2 to 0 V in steps of 0.05 V; thus giving a total of 533 input–output patterns. In our experience, because of the small number of measurements and of the low level of noise in the samples, the partitioning of this particular data set into training and testing sets does not necessarily improve the network's generalization performance. Therefore, the whole measurement data set is used for training.

In the first approach, we use an individual network to model each output of this model. This approach leads to the results presented in Table I. The accuracy of the models is measured in terms of the signal-to-noise ratio (SNR), in decibels, for each scalar output. It can be seen that the SPWL and the GRBF model provide better results than the MLP in almost all of the functions to be modeled. However, there is not a network that provides the best results globally. This is due to the different nature of the basis functions of each network. Depending on the shape of the function to be modeled, a different basis function is more suitable. Therefore, an obvious solution would be to use a mixed model combining these two kinds of networks, selecting for each output the network providing the best results. This solution, although it provides good results, presents a relatively high number of parameters to be implemented in a simulator. In order to reduce the number of parameters, a second approach consists in using a common network to perform the whole mapping globally. Table II presents the results obtained with such a solution. It can be seen that, in this case, the SPWL model provides slightly better results than the GRBF. These results are slightly lower than those provided by using a different network for each output, but the saving in the number of parameters is relevant. In general terms, the second approach is preferable: it provides a reasonable solution with a reduced number of parameters, especially when compared with conventional approaches such as look-up tables. For instance, Fig. 2 shows the intermodulation contours obtained with the model GRBF(8), with 114 parameters (a look-up table would need, in this case, 533×10 parameters), and compares them with those obtained from ex-

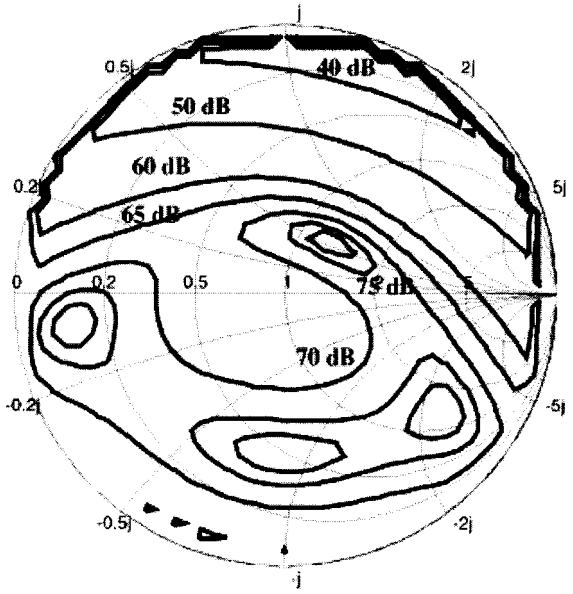

Fig. 2. Contours of the C/I ratio of the NE72084 MESFET: (a) using the measured parameters and (b) using the GRBF(8) network model.

perimental measurements. From the figure, the fidelity of the intermodulation prediction can be seen.

B. Large-Signal Results

For the large-signal model, we dispose of a set of 4050 measurements arranged in a grid of the four input variables ($V_{\rm ds}$, $V_{\rm gs}$, $v_{\rm ds}$, and $v_{\rm gs}$). A set of 1000 samples randomly selected was employed for training, and another set of 1000 samples was used as a test set to avoid overtraining. The final results are evaluated over the whole data set.

The results obtained using the different neural models are shown in Table III. The results presented correspond to solutions with a low number of parameters to facilitate the implementation of the solution in simulators. In this case, the SPWL model clearly provides the best results. The GRBF network, on the other hand, provides results similar to the MLP. As an example, Fig. 3 compares the $I-V$ measured characteristic and the approximation obtained with the SPWL with 12 neurons for a bias point of $V_{\rm ds} = 2$ V and $V_{\rm gs} = -2$ V.

TABLE III
LARGE-SIGNAL RESULTS. THE NUMBER BETWEEN PARENTHESES INDICATES THE NUMBER OF BASIS FUNCTIONS EMPLOYED FOR EACH MODEL

Model	Parameters	SNR (dB)
MLP (5)	31	24.60
GRBF (4)	36	24.95
SPWL (5)	31	28.70
MLP (9)	55	26.44
GRBF (6)	54	27.55
SPWL (10)	56	31.58
MLP (11)	67	27.36
GRBF (7)	63	28.00
SPWL (12)	66	32.67

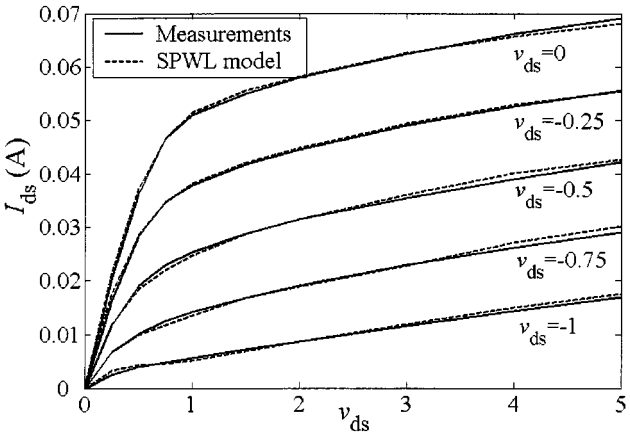

Fig. 3. Experimentally measured MESFET characteristic ($V_{\rm gs} = -2$ V, $V_{\rm ds} = 2$ V) (continuous line) and SPWL modeled surface (dashed line).

TABLE IV
TRAINING TIME FOR THE DIFFERENT MODELS

Large-signal		Small-signal	
Model	Time (min)	Model	Time (min)
MLP(11)	90	MLP(8)	47
GRBF(7)	32	GRBF(8)	15
SPWL(12)	14	SPWL(7)	7

C. Computational Burden of the Models

Another interesting aspect to be considered is the computational burden associated to each model. Table IV shows a comparison of this cost, presenting the training time of each model under the same training conditions. It can be seen that, with respect to the computational burden, the SPWL network presents the fastest training, and the GRBF is faster than the MLP network. Moreover, it can be taken into account that all these networks present the problem of local minima, and several simulations have to be performed to obtain a suitable solution, which increases the penalty for a slow training.

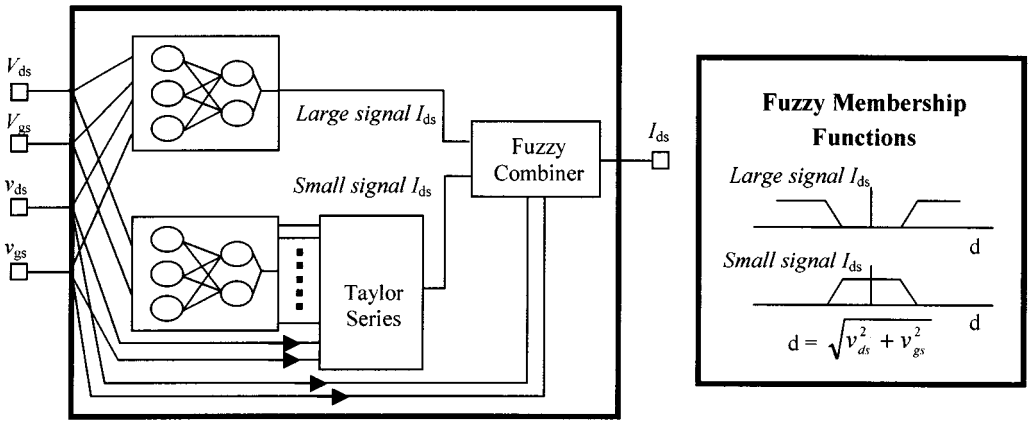

Fig. 4. Modular neural network structure for global modeling of MESFET/HEMT transistors.

V. Global Model Proposed

In the previous sections, we have treated the large- and small-signal problems of modeling a transistor separately. We have obtained a model for each regime. Therefore, one model must be selected in function of the kind of regime in which the transistor will work.

In order to avoid this unrealistic situation, we propose to use a single global model to completely characterize the transistor's behavior. This model combines the two best submodels previously obtained: one for the large-signal regime, and another one for the small-signal regime. The global model is obtained by weighting the outputs of the two submodels (modules) in a reasonable way. For example, it is possible to use a simple fuzzy combiner (Fig. 4 shows this approach). The fuzzy combiner weights each module taking into account the distance, d, of the instantaneous voltages with respect to the bias point

$$d = \sqrt{v_{\text{ds}}^2 + v_{\text{gs}}^2} \tag{21}$$

and using this distance, the membership function for the small-signal module is given by

$$\mu_{SS}(d) = \begin{cases} 1, & |d| \le d_1 \\ \dfrac{d_2 - d}{d_2 - d_1}, & d_1 < d < d_2 \\ 0, & |d| \ge d_2 \end{cases} \tag{22}$$

whereas for the large-signal regime, we have $\mu_{LS}(d) = 1 - \mu_{SS}(d)$. The parameters d_1 and d_2 defining the transition points between the two modules take in this case the values $d_1 = 0.25$ and $d_2 = 0.3$ V. These values allow an appropriate smooth transition between the large- and the small-signal modules. This global solution provides a single model that would be able to adequately characterize the whole behavior of a transistor.

VI. Conclusion

In this paper, we have presented a comparative study of several neural network solutions for the large- and small-signal modeling of MESFET and HEMT transistors.

For the small-signal regime, when using a single network, the SPWL model provides slightly better results than the GRBF network. However, when a very high accuracy is needed, the option of a mixed model, using independent SPWL and GRBF networks for each output, can be employed, paying the price of a higher number of parameters. For the large-signal behavior, the SPWL model provides clearly the best results.

Relative to the computational burden, the SPWL has proved to be the more efficient network, and the GRBF network, in any case, requires a lower computational burden than the MLP.

Finally, we have proposed a global model combining in a simple way the submodels obtained for both the large- and small-signal regimes. This model allows a whole characterization of the device, avoiding the need of working with several models for the same device.

References

[1] A. M. Crosmun and S. Maas, "Minimization of intermodulation distortion in GaAs MESFET small-signal amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. 37, pp. 1411–1417, Sept. 1989.

[2] T. Fernández *et al.*, "Extracting a bias-dependent large-signal MESFET model from $I-V$ measurements," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 372–378, Mar. 1986.

[3] W. R. Curtice and M. Ettemberg, "A nonlinear GaAs FET model for use in the design of output circuits for power amplifiers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-33, pp. 1383–1394, June 1985.

[4] A. McCamant, G. McCormak, and D. Smith, "An improved GaAs MESFET for SPICE," *IEEE Trans Microwave Theory Tech.*, vol. 38, pp. 822–824, June 1990.

[5] 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.

[6] 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, 1991, pp. 923–927.

[7] J. Rousset *et al.*, "An accurate neural network model of FET for intermodulation and power analysis," in *Proc. 26th Europ. Microwave Conf.*, Prague, Czechoslovakia, 1996.

[8] K. Shirakawa *et al.*, "A large signal characterization of an HEMT using a multilayered neural network," *IEEE Trans. Microwave Theory Tech.*, vol. 45, pp. 1630–1633, Sept. 1997.

[9] I. Santamaría *et al.*, "A nonlinear MESFET model for intermodulation analysis using a generalized radial basis function network," *Neurocomputing*, vol. 25, pp. 1–18, 1999.

[10] M. Lázaro, I. Santamaría, C. Pantaleón, A. Mediavilla, A. Tazón, and T. Fernández, "Smoothing the canonical piecewise linear model: An efficient and derivable large-signal model for MESFET/HEMT transistors," *IEEE Trans. Circuits Syst. I*, vol. 48, pp. 184–192, Feb. 2001.

[11] S. Chen, C. F. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.

[12] L. O. Chua and A. C. Deng, "Canonical piecewise-linear modeling," *IEEE Trans. Circuits Syst.*, vol. CT-33, pp. 511–525, May 1986.

[13] R. Batruni, "A multilayer neural network with piecewise-linear structure and back-propagation learning," *IEEE Trans. Neural Networks*, vol. 2, pp. 395–403, May 1991.

Marcelino Lázaro (S'00) was born in Carriazo, Spain, in 1972. He received the Telecommunication Engineer degree from the Universidad de Cantabria, Santander, Spain, in 1996, where he is currently pursuing the Ph.D. degree in the Departamento de Ingeniería de Comunicaciones. His research interests include digital signal processing, nonlinear modeling, and neural networks.

Ignacio Santamaría (M'92) was born in Vitoria, Spain, in 1967. He received the Telecommunication Engineer degree and the Ph.D. degree from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1991 and 1995, respectively.

In 1992, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, nonlinear systems, and neural networks.

Carlos Pantaleón (S'93–M'95) was born in Badajoz, Spain, in 1966. He received the Telecommunication Engineer degree and the Ph.D. degree from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1990 and 1994, respectively.

In 1990, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, nonlinear systems, and neural networks.