

# A modular neural network for global modeling of microwave transistors<sup>1</sup>

M. Lázaro, I. Santamaría, C. Pantaleón, C. Navarro, A. Tazón, T. Fernández

DICOM, ETSII y Telecom, University of Cantabria  
Avda. Los Castros, 39005, Santander, Spain  
Phone: +34-942-201392 Ext-15, Fax: +34-942-201488  
e-mail: [marce@gtas.dicom.unican.es](mailto:marce@gtas.dicom.unican.es)

## Abstract

In this paper we present a modular neural network structure for global modeling of microwave transistors (MESFET/HEMT). The model is able to accurately represent both, the small-signal and the large-signal behavior of the device. This is achieved by means of an original neural architecture, which is composed of two main modules. The first module captures the nonlinear dynamic I/V characteristic of the transistor, which governs the large signal behavior of the device. The second module estimates the derivatives at the operation (bias) point by means of a neural network and then it locally reconstructs the function by means of a third order Taylor series around that point. This second module is able to reproduce the small-signal intermodulation behavior. These two modules are combined into a global model by means of a simple fuzzy controller. In this way the global model represents adequately the device behavior independently of the nature of the applied signals.

## I. Introduction

The design of microwave and millimeter-wave circuits and the increasing integration of hybrid and monolithic circuits has reinforced the need of accurate large and small-signal device models to improve the performance of these circuits. Therefore, it is very important for efficient CAD tools to have good modeling approaches able to predict the small and large-signal nonlinear dynamic behavior of GaAs devices, such as metal semiconductor field effect transistor, (MESFET) or high electron mobility transistor (HEMT).

The general problem of modeling a microwave transistor can be stated as follows. In a transistor, the predominant nonlinear element is the drain-to-source current  $I_{ds}$ , which depends on both the drain-to-source and gate-to-source bias point ( $V_{dso}$ ,  $V_{gso}$ ) and the drain-to-source and gate-to-source dynamic voltages over the bias point ( $v_{ds}$ ,  $v_{gs}$ ). The instantaneous voltages would be the sum of both voltages, that is,  $V_{ds} = V_{dso} + v_{ds}$  and  $V_{gs} = V_{gso} + v_{gs}$ . With these premises, our modeling problem consist in finding a function  $I_{ds} = f(V_{dso}, V_{gso}, v_{ds}, v_{gs})$  that provides the estimate of the drain-to-source current as a function of the bias and the dynamic voltages.

Depending on the level of the dynamic voltages there are two clearly different regimes of behavior: the large-signal and the small-signal regimes. To model the large-signal behavior, it is enough to accurately characterize the nonlinear I/V characteristic, i.e., the dependence of  $I_{ds}$  with respect to the bias and the dynamic voltages [1]. But in a small-signal situation, to be able to model the intermodulation behavior it is necessary a different level of detail. As it is shown in [2], the nth-order intermodulation output power varies fundamentally as the square of the nth derivative of the I/V characteristic. 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, we can approximate  $I_{ds}$  by the following truncated Taylor series expansion

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

where  $I_{dso}$  is the dc drain current and ( $G_m$ , ...,  $G_{d3}$ ) are coefficients related to the nth-order derivatives of the I/V characteristic with respect to the instantaneous voltages evaluated at the bias point. Therefore, our small-signal modeling problem consists of fitting a function (model)  $g: \hat{\mathbf{A}}^2 \rightarrow \hat{\mathbf{A}}^{10}$ , which approximates the nonlinear mapping from the input space of bias voltages  $\mathbf{V} = (V_{dso}, V_{gso})$  to the output space of coefficients of the Taylor expansion  $g(\mathbf{V}) = (I_{dso}, G_m, G_{ds}, G_{m2}, G_{md}, G_{d2}, G_{m3}, G_{m2d}, G_{md2}, G_{d3})$ . Once this model is available, the drain current will be reconstructed by using the truncated Taylor series expansion (1).

---

<sup>1</sup> This work has been partially supported by CYCIT grant 1FD97-1863-C02-01

Usually, these two regimes are treated separately. In particular, we have recently proposed two different neural network structures to solve the both modeling problems: a smoothed piecewise-linear (SPWL) structure is used to model the large-signal behavior [3], and a generalized radial basis function (GRBF) network is used to estimate the function derivatives at the bias point, in order to characterize the small-signal behavior [4]. In this paper we combine these two modules into a single global model by means of fuzzy membership functions. In this way, the final global network provides a smooth transition between both regimes of behavior.

The paper is organized as follows. In Section II and III we describe the neural network modules used to model the small-signal and large-signal behavior, respectively. In Section IV we present the global model obtained from the combination of the previous models and in Section V we present the results obtained. Finally, our conclusions are presented in Section VI.

## II. A neural network model for small-signal modeling

To obtain the small-signal mapping described above, we have applied a generalized radial basis function (GRBF) network [4]. It consists of an extension of the well-known RBF network that relaxes the radial constraint of the Gaussian kernels, allowing different variances for each dimension of the input space and thus leading to elliptic basis kernels. In this way it is possible to reduce the number of basis functions and, therefore, the number of parameters. The GRBF network seems specially suited in this application because of the shape of the coefficients  $G_m \dots G_{d3}$ : while the dependence with  $V_{dso}$  is quasi-linear, the dependence with  $V_{gso}$  suggests that they could be approximated by a combination of Gaussians. The output of the GRBF network is given by

$$g(\mathbf{V}) = \sum_{i=1}^I g_i(\mathbf{V}) \quad (2)$$

where  $i$  indexes the GRBF units,  $g_i(\mathbf{V}) = \mathbf{I}_i o_i(\mathbf{V})$  and  $o_i(\mathbf{V})$  is the activation function of each unit

$$o_i(\mathbf{V}) = \prod_{j=1}^J \exp - \frac{(V_j - \mathbf{m}_{i,j})^2}{2\mathbf{s}_{i,j}^2} \quad (3)$$

where  $V_j$  is the  $j$ -th element of input vector  $\mathbf{V}$ . To train the network we have used a novel algorithm based on the Expectation-Maximization (EM) algorithm [5]. We can write the function to be fitted as  $y(\mathbf{V}) = g(\mathbf{V}) + e$ , where  $e$  is the error of the approximation, which can be assumed to be zero mean white Gaussian noise. Following the ideas expressed in [5,6], the observations  $y(\mathbf{V})$  can be decomposed into its signal and noise components

$$y_i(\mathbf{V}) = g_i(\mathbf{V}) + e_i \quad (4)$$

where the residuals  $e_i$  are obtained by decomposing the total error  $e$  into  $I$  components  $e_i = t_i e$ , and the decoupling variables  $t_i$  are restricted to sum the unity. Using this decomposition, the EM algorithm can be described as

$$\begin{aligned} E \text{ step :} & \quad \text{for } i = 1, \dots, I \text{ compute} & y_i(\mathbf{V}) = g_i(\mathbf{V}) + t_i e \\ M \text{ step :} & \quad \text{for } i = 1, \dots, I \text{ compute} & \min_{(\mathbf{?}; \mathbf{m}_{i,j}; \mathbf{s}_{i,j})} \sum_k (y_i(\mathbf{V}_k) - g_i(\mathbf{V}_k))^2 \end{aligned}$$

where  $k$  indexes the data points available. The M step is performed by means of a gradient-based method. By using this EM procedure the original complicated multiparameter optimization problem is decomposed into a set of more simple problems, which consists of estimating the parameters of each GRBF unit separately.

In [6], the decoupling variables  $t_i$  are arbitrary (but constrained to sum one); however, as it is shown in [5], an improved performance can be achieved by using as decoupling variables the posterior probabilities of the respective unit given the data.

This network provides a useful small-signal transistor model, which is able to reproduce the intermodulation distortion behavior. However, it has a clear local nature and when the dynamic voltages ( $v_{ds}, v_{gs}$ ) are large the Taylor series expansion has a loss of accuracy. In this case it is necessary to look for a large-signal model.

### III. A neural network model for large-signal modeling

Now, in the case of large dynamic signal, our modeling problem consists of obtaining a function  $G: \hat{\mathbf{A}}^4 \rightarrow \hat{\mathbf{A}}$  which approximates the nonlinear mapping from the input space  $\mathbf{V} = (V_{ds}, V_{gs}, v_{ds}, v_{gs})$  of bias and dynamic pulsed voltages to the output space, which, in this case, is directly the drain-to-source current. To carry out this nonlinear mapping we have used a smoothed piecewise linear (SPWL) model [3], which is given by

$$I_{ds}^{LS}(\mathbf{V}) = \mathbf{a} + \mathbf{B}\mathbf{V} + \sum_{i=1}^N \mathbf{c}_i \frac{1}{g} \ln(\cosh(\langle \mathbf{a}_i, \mathbf{V} \rangle - \beta_i)) \quad (5)$$

where  $\mathbf{V}$  and  $\mathbf{a}_i$  are vectors of the same dimension,  $M$ , as the input space;  $\mathbf{a}$  and  $\mathbf{c}_i$  are vectors of the same dimension of the output space,  $N$ ;  $\mathbf{B}$  is an  $N \times M$  matrix,  $\beta_i$  is a scalar,  $\langle \cdot \rangle$  denotes the inner product and  $g$  is another scalar that controls the smoothness of the model. This model is an extension of the well-known canonical piecewise linear model proposed by Chua [7], which smoothes the transition between linear regions by means of the function  $f(x) = \ln(\cosh(gx))/g$ .

The training process consists of an iterative method that first moves the partition of the input space (given by  $\mathbf{a}_i$  and  $\beta_i$ ) applying a gradient-based algorithm, and then estimates the optimal coefficients  $\mathbf{a}$ ,  $\mathbf{B}$  and  $\mathbf{c}_i$ . This model yields a smooth and derivable approximation with a low number of parameters and a reduced computational burden [3].

With this model we obtain an accurate model of the large-signal behavior of the device, but it fails when it is applied to small-signal analysis, because this model does not fit accurately enough up to the third order derivatives of the characteristic function of the device.

### IV. Proposed modular neural network

Previously we have presented two modules that characterize adequately the transistor behavior in two clearly different situations. Now, we are interested in providing a single global model capable of representing the whole transistor behavior. A simple alternative could be to combine the large and small-signal modules into a single model as it is shown in Figure 1. The two modules are combined by means of a simple fuzzy combiner that weights each module taking into account the distance,  $d$ , of the instantaneous voltages with respect to the bias point

$$d = \sqrt{v_{ds}^2 + v_{gs}^2}, \quad (6)$$

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

$$\mathbf{m}_{ss}(d) = \begin{cases} 1, & |d| \leq d_1 \\ \frac{d_2 - d}{d_2 - d_1}, & d_1 < d < d_2 \\ 0, & |d| \geq d_2 \end{cases} \quad (7)$$

whereas for the large-signal regime we have  $\mathbf{m}_{LS}(d) = 1 - \mathbf{m}_{ss}(d)$ . In (7),  $d_1$  and  $d_2$  are fixed parameters.

Conceptually it can be seen as follows. When working in large-signal regime, we need an estimate of the nonlinear dynamic  $I/V$  characteristic, and the large signal module provides it. But when working in a small-signal regime, we need a different level of detail of the device function. Then the characteristic function is locally reconstructed by using the information of the derivatives to be able to take into account the intermodulation behavior.

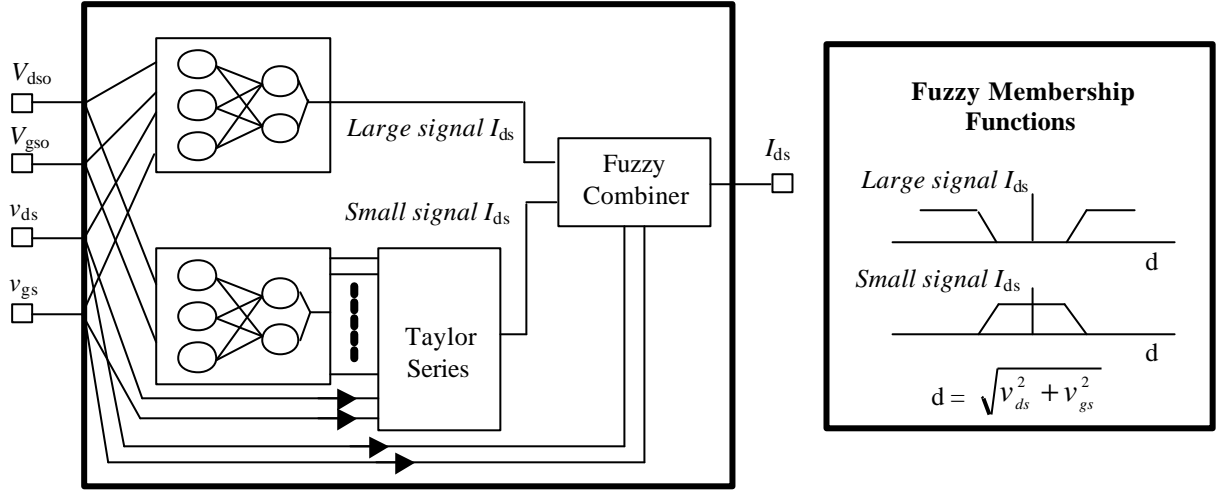


Fig. 1: Modular neural network structure for global modeling of MESFET/HEMT transistors

## V. Results

We have applied the network described above to the modeling of a MESFET transistor. The data used to train and test the model were obtained from an analytical model [8] developed from a deep study of the particular behavior of a NE72084 MESFET. The large-signal module has been trained using 12 basis functions (hyperplanes), which implies 66 parameters. The smoothing parameter,  $g$  has been trained from the information of the derivative with respect to  $V_{gs}$ , because most of the nonlinear behavior occurs along that direction. The small-signal module has been trained using 8 basis functions (gaussians), which implies 114 parameters. The fuzzy combiner was set to make the linear transition between modules between 0.25 and 0.3 V of distance with respect to the bias point: that is,  $d_1 = 0.25$  and  $d_2 = 0.3$  in (7).

Figure 2 shows the  $I/V$  characteristic function for a bias point of ( $V_{ds} = 3.5$  V,  $V_{gs} = -1$  V) and the approximation provided by the global model.

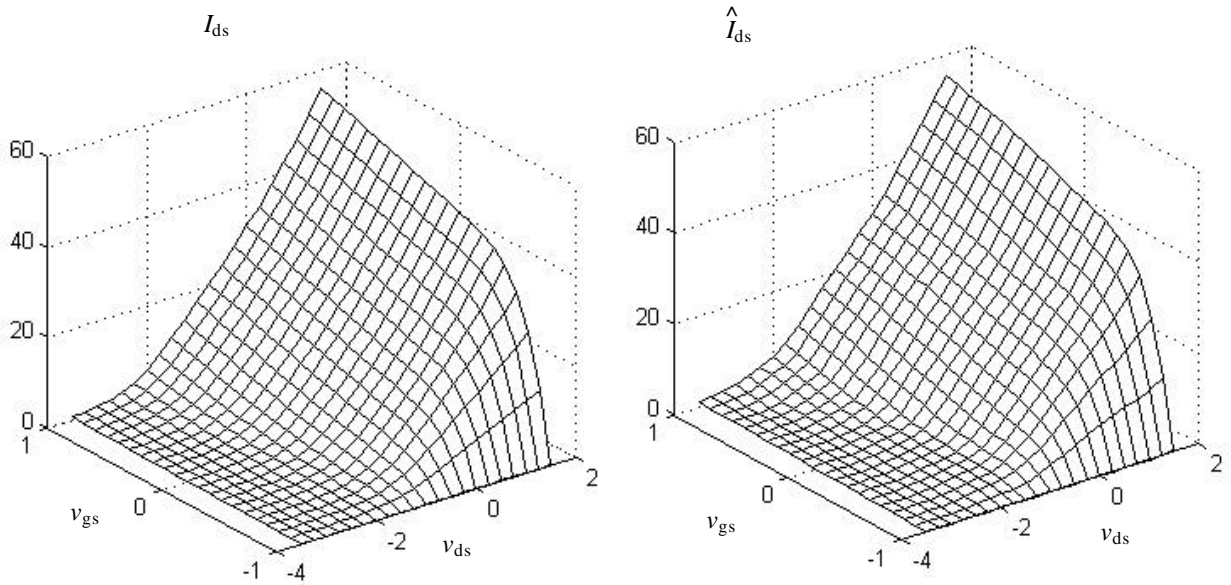


Fig.2: Characteristic  $I/V$  for a bias point ( $V_{ds} = 3.5$  V,  $V_{gs} = -1$  V). Original (left) and approximation (right).

In Figure 3 we show the behavior in a small-signal situation. Figure 3 a) shows the parameter  $G_m$  (derivative with respect to  $V_{gs} = V_{gso} + v_{gs}$  at the bias point) as a function of the bias point, while Figure 3 b) shows the corresponding estimate given by the modular model. Figure 3 c) represents the approximation of the derivative provided by the large-signal module alone: the result obtained is clearly worse than that given by the modular model. Besides, it must be noticed that  $G_m$  is the first derivative; for higher order derivatives we observe a stronger degradation. Results obtained using other typical neural network architectures (with an equivalent number of parameters), such as a multilayer perceptron, suggest that a single network can not capture the information needed to accurately model both the large and small-signal behaviors.

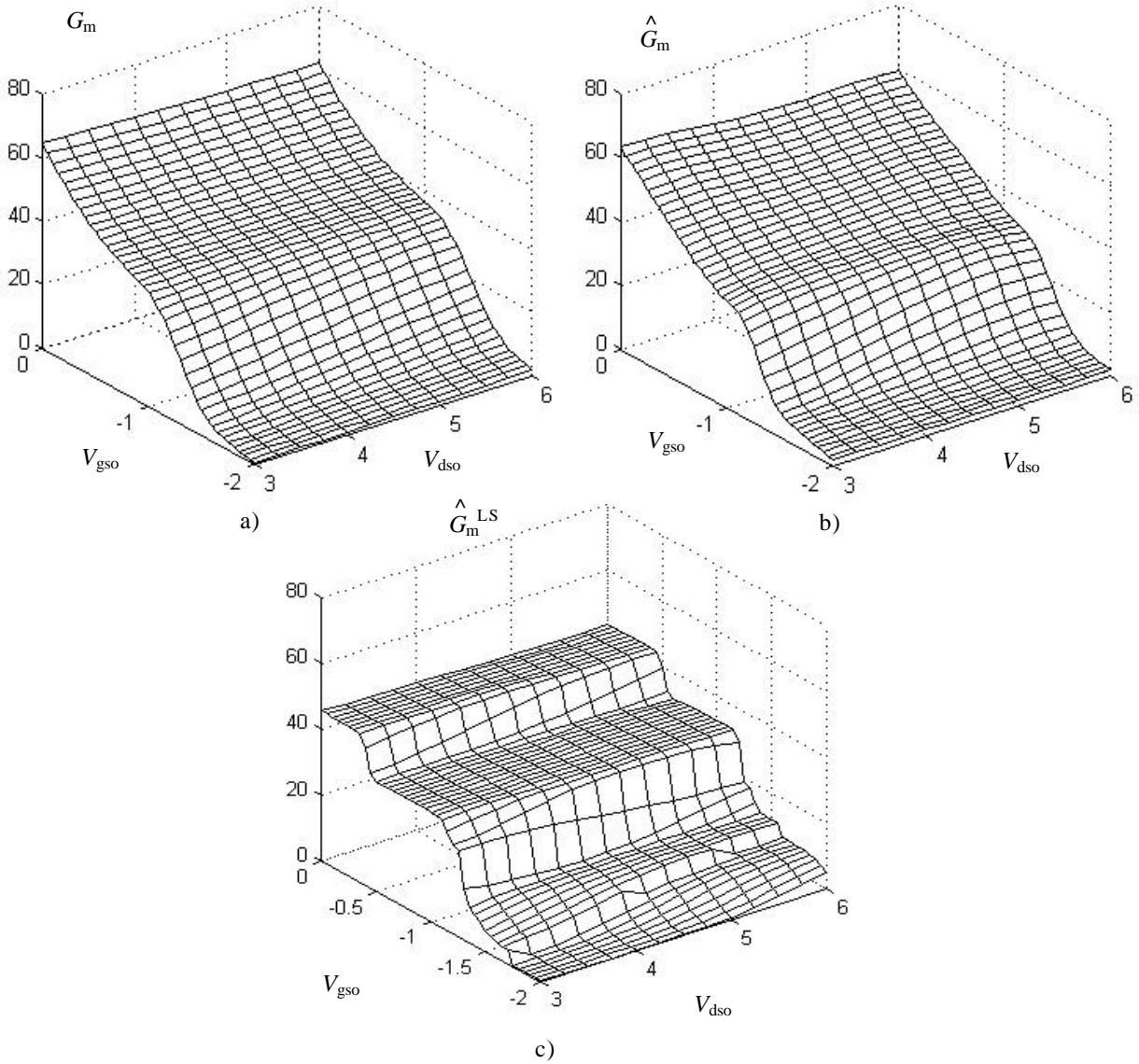


Fig. 3: Parameter  $G_m$ , a) original, b) approximation provided by the proposed modular model and c) using only the large-signal model.

To give a numerical idea of the results achieved, Table I presents the signal to noise ratio in dB reached for each of the coefficients involved in the Taylor series expansion.

| $I_{ds}$ | $G_{d2}$ | $G_{d3}$ | $G_{ds}$ | $G_m$ | $G_{m2}$ | $G_{m2d}$ | $G_{m3}$ | $G_{md}$ | $G_{md2}$ |
|----------|----------|----------|----------|-------|----------|-----------|----------|----------|-----------|
| 31.73    | 26.0     | 15.52    | 30.03    | 30.34 | 27.78    | 24.70     | 21.48    | 27.26    | 25.48     |

Table I: Results obtained in the approximation of small signal coefficients (SNR in dB)

It can be seen that these results are enough to provide a good approximation a small-signal situation. The function can be accurately reconstructed from these coefficients using the Taylor series expansion. Therefore, the proposed modular model reproduces adequately the large-signal as well as the small-signal transistor's behavior.

## VI. Conclusions

A new neural network structure has been presented that performs a global modeling of microwave transistors. It provides an accurate approximation of the whole behavior of the device combining two modules that capture a different kind of behavior. A first module is responsible of capturing the nonlinear dynamic  $I/V$  characteristic of the device, which drives the large signal behavior. The other module is responsible of the local reconstruction of the  $I/V$  characteristic, taking into account the information of the derivatives, in order to represent the small-signal intermodulation behavior. The global model presents a reduced number of parameters, and the computational burden to carry out the training process is lower than that required by other networks, like the MLP, which allows an easy implementation in practical simulators.

## References

- [1] T. Fernández et al. "Extracting a bias-dependent large signal MESFET model from pulsed  $I/V$  measurements". *IEEE Trans. Microwave Theory Tech.*, vol 44, no. 3, pp. 372-378, 1996.
- [2] A. M. Crosmun, S. Maas, "Minimization of intermodulation distortion in GaAs MESFET small-signal amplifiers", *IEEE Trans. Microwave Theory Tech.*, vol. 37, no. 9, pp. 1411-1417, 1989.
- [3] M. Lázaro, I. Santamaría, C. Pantaleón et al. "Smoothing the canonical piecewise linear model: an efficient and derivable large-signal model for MESFET/HEMT transistors", Submitted to *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*.
- [4] I. Santamaría, M. Lázaro, C. Pantaleón, J. García, A. Tazón, A. Mediavilla, "A nonlinear MESFET model for intermodulation analysis using a generalized radial basis function network", *Neurocomputing*, vol. 25, pp. 1-18, 1999.
- [5] I. Santamaría, M. Lázaro, C. Pantaleón, "EM-based training of GRBF networks with application to nonlinear MESFET modeling" submitted to ICASSP2000.
- [6] M. Feder, E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 477-489, 1988.
- [7] L. O. Chua, A. C. Deng, "Canonical piecewise-linear modeling", *IEEE Trans. Circuits Syst.*, vol. 33, no. 5, pp. 511-525, 1986.
- [8] C. Navarro et al, "Large signal dynamic properties of GaAs MESFET/HEMT devices under optical illumination", *Proc. of the GAAS'98 Symposium*, pp. 350-353, Amsterdam, 1998.