Smoothing the Canonical Piecewise-Linear Model: An Efficient and Derivable Large-Signal Model for MESFET/HEMT Transistors

Marcelino Lázaro, Ignacio Santamaría, Carlos Pantaleón, Angel Mediavilla Sánchez, *Member, IEEE*, Antonio Tazón Puente, *Member, IEEE*, and Tomás Fernández

***Abstract*—In this paper we present the smoothed piecewise-linear (SPWL) model as a useful tool in the device modeling field. The SPWL model is an extension of the well-known canonical piecewise-linear model proposed by Chua, which substitutes the abrupt absolute value function for a smoothing function (the logarithm of hyperbolic cosine), thus providing the model with several interesting properties. In particular, this function makes the model derivable, which is important to predict the intermodulation distortion behavior. Moreover, it allows one to control the smoothness of the global model by means of a single smoothing parameter. The parameters of the model are adapted to fit the nonlinear function, while the smoothing parameter is selected according to derivative constraints. The applied learning algorithm is a second-order gradient method. The proposed SPWL model is successfully applied to model a microwave HEMT transistor under optical illumination using real measurements. The model receives as input the bias voltages of the transistor, the instantaneous voltages, and the optical power and provides the drain to source current. The performance and computational burden of the SPWL model is compared with an empirical model and with some neural networks-based alternatives.**

***Index Terms*—MESFET/HEMT modeling, nonlinear modeling, piecewise-linear modeling, smoothing methods.**

I. Introduction

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

Conventional nonlinear techniques applied to device modeling, such as closed-form equations [1], [2], Volterra series [3], or the use of look-up tables [4], are difficult to implement in commercial simulators because of their high memory requirements or their computational burden. Moreover, these techniques fail to reproduce adequately the nonlinear I/V function derivatives around the bias point, which is interesting, for instance, to model and predict the intermodulation distortion behavior [5]. Most of the MESFET's and HEMT's models used for years were not conceived for intermodulation prediction and have poor performance when this nonlinear behavior either in large-signal or small-signal regimes is of primary interest.

Recently, some attempts have been made to model the nonlinear behavior of active devices and circuits by using neural networks [6]–[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 of these neural approaches only consider the use of the multilayer perceptron (MLP) and, in this case, the memory requirements to give an accurate approximation, and the computational requirements to carry out the training process of the network, are high. In some specific applications, for instance, to model the derivatives around the bias point for small-signal intermodulation prediction, a different neural architecture, the generalized radial basis function (GRBF) network [9], [10], has shown better performance than the MLP. However, the semilocal activation function used in the GRBF network is not adequate for large-signal modeling problems. Besides, like all the radial basis function based networks, the GRBF requires a large number of units when the input space dimensionality is high.

An interesting alternative, specially suited for nonlinear device modeling, is the canonical piecewise-linear (CPWL) model proposed by Chua [11], [12]. This model provides accurate approximations with a low number of parameters and with a computational burden lower than the neural networks solutions. However, it lacks the capability of approximating the derivatives of the function because of its piecewise-linear nature: the second and higher order derivatives are always zero. In [13], a generalization of the CPWL model is proposed, which substitutes the local linear activation functions for polynomial functions, thus yielding a piecewise-smooth model. Although the higher derivatives for this piecewise-smooth model are no longer zero, the use of polynomial functions does not seem appropriate for device modeling problems, since they tend to oscillate when the input space is partitioned into just a few large regions.

In this paper, we propose a different extension of the canonical piecewise-linear model, which, retaining the advantages

Manuscript received October 15, 1999; revised June 10, 2000. This work has been supported in part by the European Commission and CICYT under Grants TIC96-0500-C10-07 and 1FD97-1863-C02-01.This paper was recommended by Associate Editor Y. Park.

The authors are with the Department Ingeniería de Comunicaciones Universidad de Cantabria, 39005, Santander, Spain (e-mail: marce@gtas.dicom.unican.es).

Publisher Item Identifier S 1057-7122(01)01391-5.

of the PWL model, makes it smooth and derivable. We denote this generalization as the smoothed piecewise-linear (SPWL) model. The SPWL uses a summation of smooth functions (in particular, it uses the logarithm of hyperbolic cosine), of the boundary functions instead of the absolute values used in the conventional CPWL model. This modification allows the model to be derivable, smoothing the transitions between different linear regions of the input space. A single parameter controls the smoothness (the integral of the squared second derivative) of the new activation function and, as it is shown in the paper, the smoothness of the global model. This fact provides a link with regularization theory; in particular, the parameter controlling the smoothness can be seen as a regularization parameter.

The paper is organized as follows. In Section II we present the SPWL model where first, we revise the canonical PWL model; second, we propose and discuss functions to smooth the model. Finally, it is shown that, by using the selected function, the smoothness of the global model is a monotone function of the parameter that controls the smoothness between boundaries. Section III describes the learning algorithm employed to obtain the parameters of the model. In Section IV the proposed SPWL model is applied to approximate the I/V nonlinear characteristic of a microwave HEMT, including the effect of the optical power illumination applied to it. Using real measurements, a SPWL model is adjusted; it receives as inputs the optical power, the bias voltages and the instantaneous voltages and provides as output the drain to source current. A comparison is made with the standard PWL model, with an empirical model based on closed-form equations as well as with some neural network alternatives. Finally, the conclusions are presented in Section V.

II. The Smoothed Piecewise-Linear Model

A. The Canonical piecewise-linear Model

Although it is known that the canonical PWL is a subset of the general class of PWL functions, it is the only one that admits an explicit and compact formulation [14]. In its basic formulation, the canonical PWL model proposed by Chua [11], [12] performs a mapping $f\colon R^M \to R^M$ as follows:

$$f(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x} + \sum_{i=1}^{\theta} \boldsymbol{c_i} |\langle \boldsymbol{\alpha_i}, \boldsymbol{x}\rangle - \beta_i| \tag{1}$$

where

$\boldsymbol{a}$, $\boldsymbol{x}$, $\boldsymbol{c_i}$, and $\boldsymbol{\alpha_i}$	vectors of the same dimension, M, as the input space;
$\boldsymbol{B}$	$M \times M$ 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 $M-1$. In each region of the domain space, the function is composed by a linear combination of linear hyperplanes. It can be seen that the expression inside the absolute value function defines the boundaries partitioning the domain space. It has the characteristic of the absolute value in zero: it is continuous but not derivable. For device modeling problems, this model is capable to provide a good performance with a low number of parameters, but it lacks the capability of approximating the derivatives.

B. Smoothing the Absolute Value Function

The smoothed PWL model that we propose in this section overcomes this lack of derivability without increasing the memory requirements. This is achieved by substituting in (1) the abrupt absolute value for a derivable function in order to smooth the joint of hyperplanes at the boundaries defining the domain space partition.

There are several possibilities to carry out smooth transitions between linear regimes. To illustrate this point, let us consider a function $f\colon R \to R$ composed by two linear sections separated by a breakpoint α

$$f(x) = \theta_0 + \theta_1(x-\alpha) + \theta_2|x-\alpha| \tag{2}$$

equivalently, (2) can be rewritten using the signum (sgn) function as follows:

$$f(x) = \theta_0 + \theta_1(x-\alpha) + \theta_2(x-\alpha)\,\mathrm{sgn}(x-\alpha). \tag{3}$$

It would be clearly desirable to have a family of functions providing a smooth transition and allowing parametric control of the "sharpness" of the transition. There are several such functions in the bibliography; for example, in [15] the $\mathrm{sgn}(x)$ is replaced by a member of a family of functions, denoted as $trn(\gamma x)$. They include a parameter, γ, that controls the smoothness of the transition; by changing γ we control the smoothness around the breakpoint. Some of these functions are [16]

$$f\tan(x,\gamma) = x\tanh(\gamma x) \tag{4}$$

and

$$f\,\mathrm{pol}(x,\gamma) = \begin{cases} \dfrac{1}{8\gamma}[15(\gamma x)^2 - 10(\gamma x)^4 + 3(\gamma x)^6], & |x| \le \dfrac{1}{\gamma} \\ |x|, & |x| > \dfrac{1}{\gamma}. \end{cases} \tag{5}$$

A geometrical parameter, which can give some insight about the behavior of these smoothing functions, is the radius of curvature, which is given by

$$K = \frac{f''(x)}{(1+(f'(x))^2)^{3/2}} \tag{6}$$

at the breakpoint $x=0$, the radius of curvature for (4) and (5) is 2γ and $15\gamma/4$, respectively. It can be seen that for high values of K, the functions tend faster to the absolute value function. Also, it can be shown that both functions reach values greater than one for the first derivative and negative values for the second derivative. Their derivatives, therefore, have "overshootings" that make difficult the task of obtaining smooth derivatives. To alleviate this problem other smoothing functions can be used; for instance, Griffiths and Miller in [17] proposed replacing the $|x|$ function by

$$\mathrm{hyp}(x,\gamma) = \sqrt{x^2+\gamma} \tag{7}$$

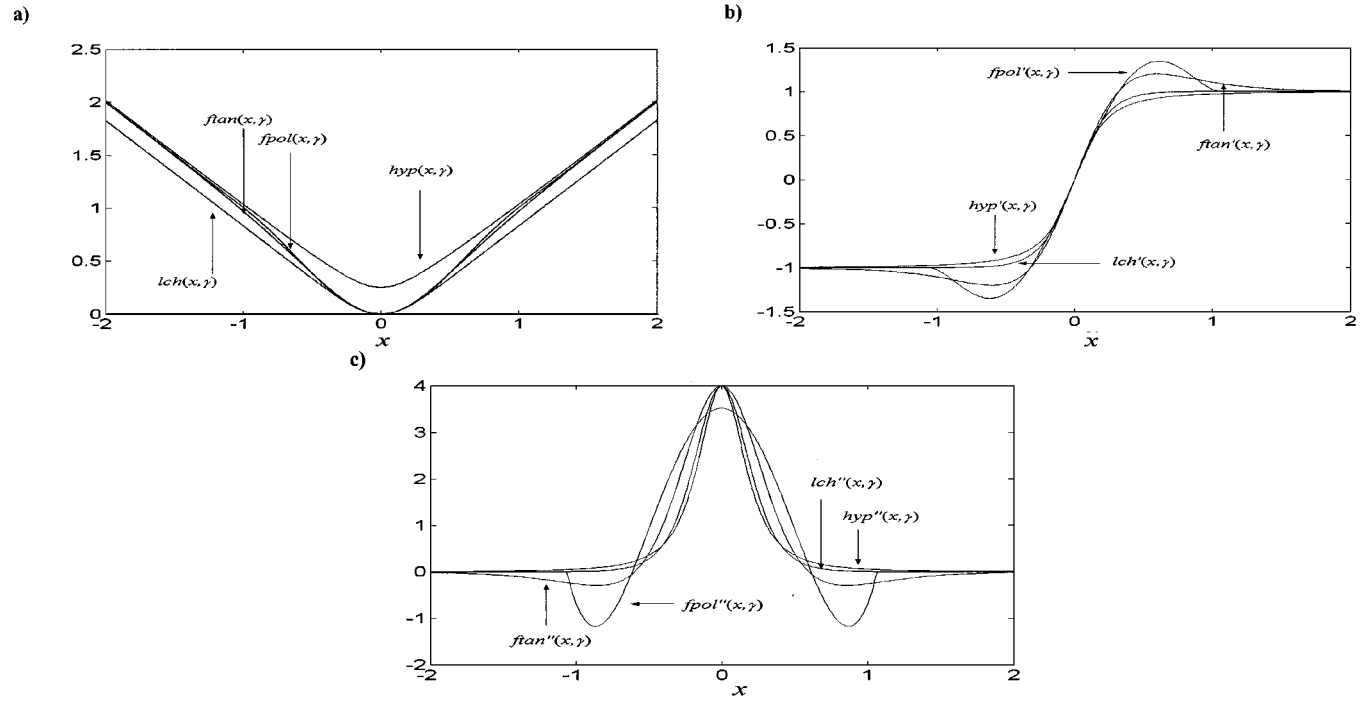

Fig. 1. Smoothing functions and their derivatives. (a) Smoothing functions. (b) First derivatives. (c) Second derivatives.

with radius of curvature $K = 1/\sqrt{\gamma}$. In this paper, we propose the function

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

with radius of curvature $K = \gamma$.

Fig. 1 shows the four smoothing functions ($f\tan(x, \gamma)$, $f\,\mathrm{pol}(x, \gamma)$, $\mathrm{hyp}(x, \gamma)$, and $\mathrm{lch}(x, \gamma)$), as well as their derivatives for a value $K = 4$. We can see that the derivatives of $\mathrm{hyp}(x, \gamma)$ and $\mathrm{lch}(x, \gamma)$ do not present overshootings; moreover, it can be proven that, for the same curvature, these two functions are smoother (i.e., the squared second derivative has smaller area) than $f\tan(x, \gamma)$ and $f\,\mathrm{pol}(x, \gamma)$. Besides, we have observed that, in practice, $\mathrm{hyp}(x, \gamma)$ and $\mathrm{lch}(x, \gamma)$ give the best performance for device modeling problems. Finally, we have chosen the function $\mathrm{lch}(x, \gamma)$ to smooth the canonical piecewise-linear model (1) mainly for notational convenience, since its derivatives can be easily calculated as $\mathrm{lch}'(x, \gamma) = \tanh(x\gamma)$ and $\mathrm{lch}''(x, \gamma) = \gamma\,\mathrm{sech}^2(x, \gamma)$.

Although in the above discussion we have considered one-dimensional (1-D) functions, the extension to an input space of higher dimensionality is straightforward. Therefore, the smoothed piecewise-linear (SPWL) model performs a mapping f: $R^M \rightarrow R^M$ as follows:

$$f(\boldsymbol{x}) = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x} + \sum_{i=1}^{\theta} \boldsymbol{c}_i \,\mathrm{lch}(\langle \boldsymbol{\alpha}_i, \boldsymbol{x}\rangle - \beta_i, \gamma_i). \tag{9}$$

Now, the ith boundary is governed by the function $\mathrm{lch}(\langle \boldsymbol{\alpha}_i, \boldsymbol{x}\rangle - \beta_i, \gamma_i)$, a smooth function, continuous and derivable, and consequently, the whole model is endowed with these properties.

C. Smoothness of the SPWL Model

In this section we show that the smoothness of the global model is a monotone function of the parameter γ (here we will consider a simplified model using the same γ for all the boundaries). Then, γ behaves like a regularization parameter, allowing a tradeoff between fidelity to the measurements and smoothness of the model.

Standard regularization techniques minimize a cost functional consisting of two terms: the first one measures the closeness to the data and the second term weights the cost associated with a functional that measures the smoothness of the solution, i.e.,

$$E = \sum_i (f(x_i) - y_i)^2 + \lambda \|Pf(x)\|^2 \tag{10}$$

where

- y_i measurements;
- λ regularization parameter, which controls the compromise between degree of smoothness of the solution and its closeness to the data;
- P functional (stabilizer).

Smoothness can be measured in a number of different ways; generally, the stabilizer P involves some derivatives of the function. A widely used class of stabilizers is given by the following functionals [18]:

$$\|P^m f\|^2 = \sum_{i_1 \ldots i_m}^{n} \int_{R^n} dx (\partial_{i_1 \ldots i_m} f(x))^2 \tag{11}$$

where $\partial_{i_1 \ldots i_m} = \partial^m / \partial x_{i_1} \ldots \partial x_{i_m}$ and $m \geq 1$. We will use here this type of stabilizers, which are invariant under rotation and translation.

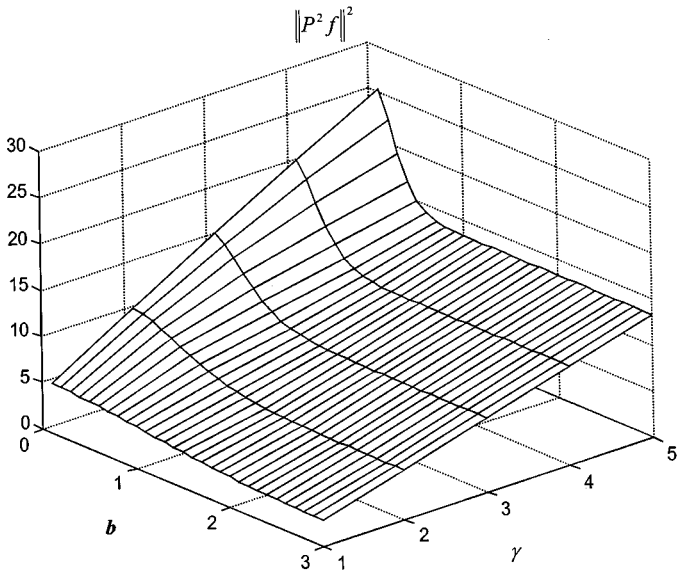

Fig. 2. Smoothness functional for the one dimensional SPWL model.

With these premises, we will point out a link between the SPWL model and regularization theory. Specifically, we will show that the relation between the smoothness (evaluated by means of integrals of the second order derivatives) and the parameter γ is monotone and that, therefore, it is possible to control the model smoothness by means of an adequate selection of γ.

Let us start with a 1-D example: in this case, the particular expression of the stabilizer is the squared second derivative of the model. On the other hand, the second derivative of each unit has a $\operatorname{sech}^2(x\gamma)$ shape, which is a localized function. Therefore we can assume that the second derivative of each component only overlaps with the nearest components. Without lack of generality, we can assume that the model is composed of two weighted components, separated a distance b

$$f(x) = c_1 \operatorname{lch}(x, \gamma) + c_2 \operatorname{lch}(x - b, \lambda). \tag{12}$$

Then, the regularization term is given by

$$\|P^2 f\|^2 = \frac{4}{3}(c_1^2 + c_2^2)\gamma + 8c_1c_2\gamma \operatorname{cosech}^2(\gamma b) \cdot \left[\frac{\gamma b}{\tanh(\gamma b)} - 1\right] \tag{13}$$

which is monotone with γ independently of b, fixing the separation between components, or of the weighting parameters c_1 and c_2. Fig. 2 shows the value of this functional versus the distance b and the smoothing parameter γ, for $c_1 = c_2 = 1$. The same behavior is observed for any value of the weighting coefficients.

To extend the above analysis to a bidimensional input space, we encounter the problem that now the second derivative has not finite support along the direction of the boundaries. However, we can avoid this problem by integrating over a suitable finite region. To simplify the evaluation of the integrals, straight lines orthogonal to the boundaries delimit the region of integration. Fig. 3 shows this kind of region for two boundaries in the input space. Now, denoting the region of integration as R', the stabilizer has the following expression:

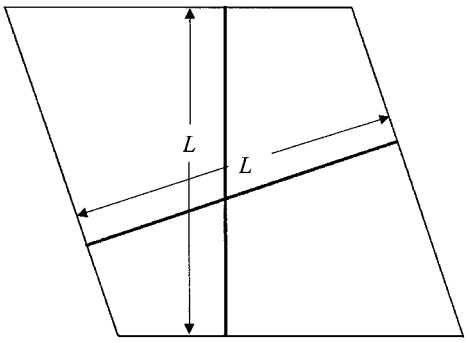

Fig. 3. Region of integration for a bidimensional input space.

$$\|P^2 f\|^2 = \int_{R'} \left[\left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2\right] dx_1\, dx_2. \tag{14}$$

Considering again a model composed of two overlapped components, we have

$$f(\boldsymbol{x}) = \sum_{i=1}^{2} c_i \operatorname{lch}(m_i x_1 - x_2 + t_i, \gamma). \tag{15}$$

Without lack of generality, we can assume that one of the components has $m_1 = t_1 = 0$, and denote the parameters of the other component as $m_2 = m$, $t_2 = t$. Then, the regularization term is calculated as

$$\begin{aligned}\|P^2 f\|^2 = &\int_{R'} c_1^2\gamma^2 \operatorname{sech}^4(-\gamma x_2)\, dx_1\, dx_2 \\ &+ \int_{R'} c_2^2\gamma^2 \operatorname{sech}^4(\gamma(mx_1 - x_2 + t))\, dx_1\, dx_2 \\ &+ \int_{R'} c_1c_2\gamma^2 \operatorname{sech}^2(-\gamma x_2) \\ &\cdot \operatorname{sech}^2(\gamma(mx_1 - x_2 + t))\, dx_1\, dx_2.\end{aligned} \tag{16}$$

The first two terms correspond to the contribution of each component, while the third term corresponds to the intersection of both components. Let us point out that the last term can be integrated over the whole input space, since the intersection of two $\operatorname{sech}^2(x)$ functions is finite. The expression finally obtained is

$$\|P^2 f\|^2 = \frac{4}{3}\gamma L(c_1^2 + c_2^2) + \frac{4}{m} c_1c_2. \tag{17}$$

Equation (17) shows again a linear and monotone dependence of the smoothness of the model with γ. To increase γ makes the transitions between linear sections more abrupt and correspondingly the smoothness of the global model is reduced.

The extension of the previous results to an input space of dimension higher than two seems rather involved and it has not been pursued here. Nevertheless, it is reasonable to assume the same behavior with respect to the smoothing parameter.

As a conclusion of this section, we have shown that the smoothing parameter γ can be seen as a standard regularization

parameter. In this way, to minimize the squared error alone, together with a proper selection of γ according to some smoothness constraints (derivative constraints, for instance) is equivalent to minimize a regularized functional as (10).

III. Model Training

The SPWL model has three different kinds of parameters: those defining the boundaries partitioning the domain space: $\boldsymbol{\alpha_i}$ and β_i; those defining the linear combination of the model components: $\boldsymbol{a}$, $\boldsymbol{B}$, and $\boldsymbol{c_i}$; and the smoothing parameter γ.

The learning algorithm of the SPWL model is an iterative algorithm based in the successive adaptation of the partition of the domain space and the estimate of the optimal coefficients defining the linear model components for that given partition. The adaptation of the parameters defining the boundaries in the domain space is based in a second-order gradient method, i.e., the gradient and the Hessian of the error function with respect to the parameters are calculated and used to carry out the adaptation. Once the boundaries of the domain space are fixed, then the approximation error is a quadratic function of the parameters defining the linear combination of the components, and the minimum can be easily found by solving a linear least squares problem. Then, the boundaries are adapted again and the process is repeated iteratively. This is basically the method proposed by Chua to optimize the parameters of the canonical piecewise-linear model [11] and it is particularized for our smoothing function in Appendix A.

Finally, we will consider the estimation of γ, which is a key parameter of the SPWL model. Specifically, we will describe its estimation for large-signal device modeling problems. Several strategies are possible: the simplest one consists of using a γ selected to minimize the error of the approximation and, at the same time, to fulfill some smoothness constraint. The optimal γ can be obtained by applying a gradient descent algorithm

$$\gamma_{k+1} = \gamma_k - \mu \frac{\partial E}{\partial \gamma} \tag{18}$$

where E is the mean squared error. A different γ can be used for each boundary; however, the improvement over using a common γ for all the boundaries does not compensate for the increase in the number of model parameters.

A more interesting alternative for device modeling problems is to use information about the function derivatives. As it was said before, to reproduce the intermodulation distortion behavior it is necessary to model the higher order derivatives. It seems reasonable, therefore, to look for a tradeoff between the approximation of the function and the approximation of the derivatives. If some information about the derivatives is available (for instance, obtained from two tones measurements), then the optimal γ parameter can be selected according to it. For example, let us assume that it is possible to measure the first derivative of the model with respect to the jth input parameter; $\boldsymbol{y'} = (y'_1, \ldots, y'_N)^T$, then γ can be selected to minimize

$$E_1 = \sum_{l=1}^{N} \left(y'_l - \left(b_j x_{j,l} + \sum_{i=1}^{\theta} c_i m_{i,j} \tanh(\gamma b_i(\boldsymbol{x}_l)) \right) \right)^2 \tag{19}$$

where the inner parentheses corresponds to the derivative of the SPWL model with respect to the jth input parameter. In this case, the optimal γ is obtained by applying

$$\gamma_{k+1} = \gamma_k + \mu \sum_{l=1}^{N} e'_l \left(\sum_{i=1}^{\theta} c_i m_{i,j} b_i(\boldsymbol{x}_l) \operatorname{sech}^2(\gamma b_i(\boldsymbol{x}_l)) \right). \tag{20}$$

Again, a different smoothing parameter can be used for each boundary.

As a conclusion of this section, we can say that one of the most relevant characteristics of the SPWL model is that we can take advantage of the additional degree of freedom provided by γ without degrading noticeably the fit to the function.

IV. Large-Signal Modeling of HEMT Transistors Under Optical Illumination

A. Background

The distinct advantages of optical transmission systems and the increasing use of microwave frequencies within general communication systems, coupled with the ability to integrate microwave and optical components onto a single slice of GaAs, have stimulated considerable interest in the development of microwave opto-electronic systems. The optical circuits are advantageous because they can be integrated into the microwave circuits without interfering with them, and they have low losses and small dimensions, short reaction time, and wide band. Direct illumination on the microwave or millimeter-wave monolithic circuit is very attractive for versatility of applications associated with the optical fiber communication and control systems. The GaAs FET, the basic building block of MMIC's, can be used as a photo-detector embedded on the monolithic chip itself, and thus serves as an optical port. Then, it is significant to examine optical-microwave interaction on a FET in a monolithic circuit and how the variation of the FET model parameters is due to the illumination.

It is well known that when we illuminate an GaAs device, an interesting absorption effect takes place at the Gate-Drain and Gate-Source spacing, and a free carrier photoexcitation occurs at the active area level. In fact these devices exhibit both photoconductive and photovoltaic effects that can be conveniently amplified by using external buffer resistors. This means that the static DC curves as well as the small signal equivalent circuit parameters change when optical energy goes into the device [19], [20].

However, the true large signal behavior is governed by the dynamic pulsed I/V characteristic that depends on the quiescent bias point. As far as we know, the only report on the effects of optical illumination on this bias dependent dynamic behavior that is responsible of the output power and efficiency capabilities of these electron devices is an analytic functions based method [21]. This method is based in an extensive investigation on the large signal dynamic behavior of the device. It studies its different dependencies and then fits these dependencies by means of a set of suitable analytic functions. For example, in the particular case of a HEMT, it has identified a logarithmic dependence with the optical power and a hyperbolic tangent shape with the

gate voltage. Although this method obtains acceptable results, it has the drawback of being specific for each device. For each new device, the whole study must be repeated and the optimal analytical functions must be selected again to fit the specific data. This kind of process is time consuming, and an automated method that could be used independently of the specific device and only from the available measured data would be desirable.

In the next section, we present the results obtained in the application of the SPWL model to this problem, as well as its comparison with other neural network alternatives.

B. Modeling Results

The experimentally measured data for a Philips D02AH ($4 * 30$ μm) microwave HEMT, with 0.2-μm gate length, were employed to train the SPWL network in order to accurately model the nonlinear dependence of the drain to source current I_{ds} with respect to the bias voltages (V_{gso}, V_{dso}), the instantaneous voltages (v_{gs}, v_{ds}), and the incident optical power (P_o). The dc behavior of the transistor was measured at the following bias points: $V_{dso} = -0.75$ V, 0 V, and 0.75 V; $V_{gso} = 0$ V, 2 V, and 4 V. For the instantaneous voltages, v_{ds} is swept from 0 V to 4 V in steps of 0.25 V, and v_{gs} from -1 V to 0.75 V in steps of 0.25 V. Finally, these measurements were repeated for the following incident optical power: $P_o = 0$ mW, 0.25 mW, 0.50 mW, 0.75 mW, 1 mW, 2 mW, 5 mW, and 10 mW. With this grid, we dispose of a set of 9792 input/output samples, which have been used to train the different models.

We adjusted and compared the performance of five different models: an MLP, a GRBF network, an analytical model [9], a conventional CPWL model, and the proposed SPWL model.

To evaluate the accuracy of each model, a signal to noise ratio (*SNR*) for the I_{ds} estimate provided by the model was employed as a figure of merit.

To perform the training of the CPWL, SPWL, and GRBF models a data training subset of 1000 randomly chosen samples was selected. The models were validated using the rest of the points and the *SNR* was evaluated using the whole data set available (9792 data points). Using this smaller training data set the computational burden of the learning process is reduced without affecting their generalization capability. For the training of the MLP model, better results were obtained by using a larger training set of 2000 points and a test set of 7292 points to avoid overtraining. The MLP was trained using the conventional backpropagation algorithm with an adaptive learning rate, while the GRBF network was trained adding a centroid at a time and retraining the whole network following the algorithm described in [9].

Table I presents the SNR values for the I_{ds} estimates and for the first derivatives with respect to the instantaneous voltages (v_{gs} and v_{ds}), provided by the different models. For the transistor used in this example the true derivatives have not been measured, instead we have fitted a cubic spline model to the whole data set (giving an unpractical model with a huge number of parameters) and then we estimated the derivatives from that spline model. In this example, the SPWL model was fitted using the algorithm described in Section III, adapting the smoothing parameter by means of the information of the first derivative. From the Table I, we can see that for a given number of parameters, the SPWL model provides the best

TABLE I
COMPARISON OF MODELING RESULTS FOR A HEMT: PHILLIPS D02AH ($4 * 30$ μm): THE FIRST COLUMN INDICATES THE MODEL AND THE SECOND THE NUMBER OF PARAMETERS OF THE MODEL. THE REMAINDER THREE COLUMNS INDICATES THE APPROXIMATION RESULTS FOR THE FUNCTION AND THE DERIVATIVES WITH RESPECT TO v_{ds} AND v_{gs} RESPECTIVELY EXPRESSED BY THE SNR RATIO IN dB

Model	Number of Parameters	Function SNR (dB)	d/dv_{ds} SNR (dB)	d/dv_{gs} SNR (dB)
MLP(5)	36	25.50	14.00	14.40
CPWL(5)	36	30.62	9.52	13.04
SPWL(5)	**37**	**32.13**	**16.63**	**16.40**
MLP(9)	64	27.17	12.62	14.75
CPWL(10)	66	32.44	10.44	12.96
SPWL(10)	**67**	**34.35**	**18.40**	**17.15**
MLP(11)	78	28.09	13.71	14.66
CPWL(12)	78	33.04	11.51	14.34
SPWL(12)	**79**	**35.16**	**18.88**	**17.16**
GRBF(8)	88	28.84	10.09	13.79
Analytical Model	98	32.92	-0.8674	16.53

MLP(N): Multilayer Perceptron with N neurons in the hidden layer.

GRBF(N): Generalized Radial Basis Function with N neurons.

Analytical Model: Analytic functions proposed in [21].

CPWL(N): Canonical Piecewise Linear Model with N boundaries in the input space.

SPWL(N): Smoothed Piecewise Linear model with N boundaries in the input space.

results for both the function and the derivatives. Fig. 4 shows the original function and derivatives and the approximation given by the model for (V_{gso}, V_{dso}), the instantaneous voltages (v_{gs}, v_{ds}), an incident optical power $P_o = 1$ mW, and a bias voltage $V_{gso} = 0$ V and $V_{dso} = 2$ V.

An important aspect is the computational burden needed to estimate the model parameters. A rough estimation, given by the training time necessary to carry out the training of the different methods, is presented in Table II. The different models have been implemented in MATLAB programs over a Pentium MMX 200 Mhz. It can be seen that the required time is significantly lower for the CPWL and the SPWL models than for the other models. While these models require a few minutes to train the model, the MLP and the GRBF need several hours. This means a great advantage for both PWL models. For the method in [21], there is more important the time consumed by the data analysis that the computational burden because it is a method based on the exhaustive investigation of the specific device in order to extract the dependencies. Moreover, it does not work in an automated way and therefore the whole process must be repeated for each new device to model.

V. CONCLUSIONS

The SPWL model has been proposed as a useful tool to model the large-signal behavior of MESFET/HEMT transistors. It al-

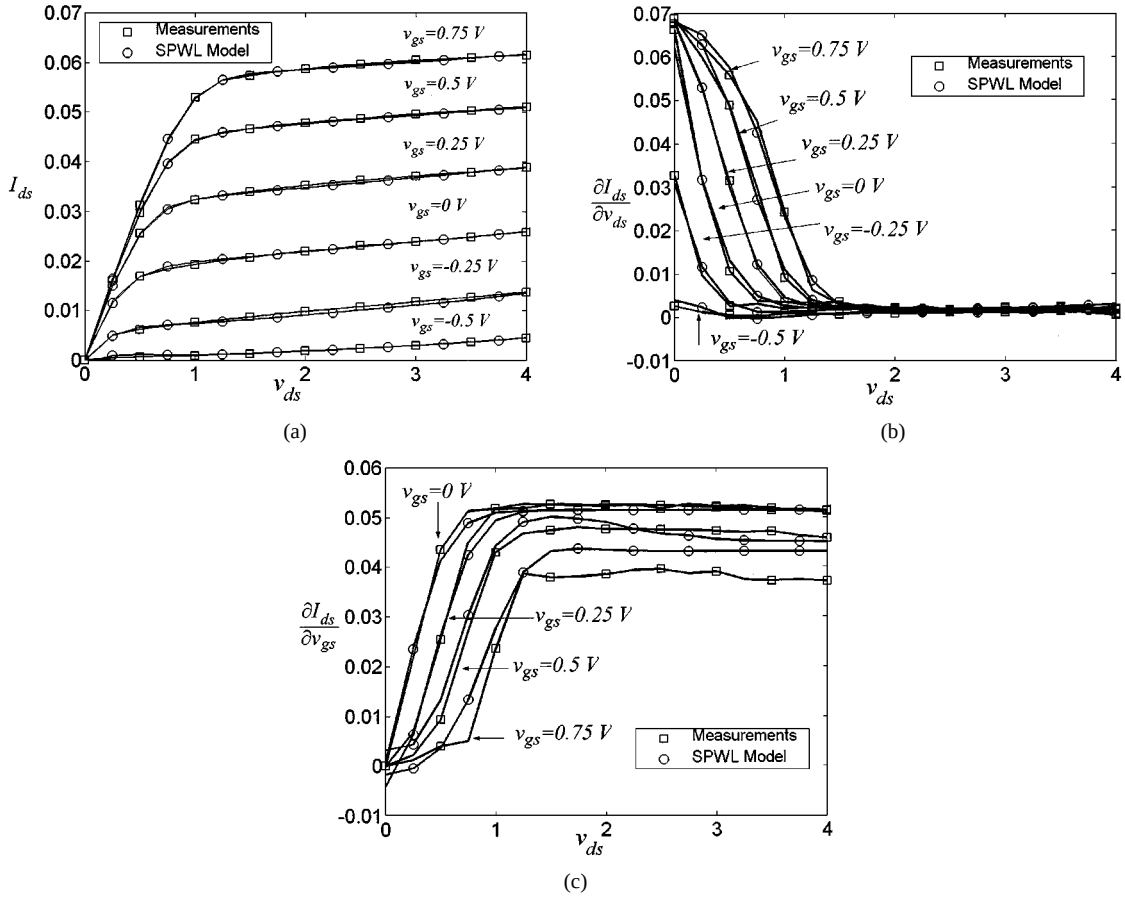

Fig. 4. Experimentally extracted and SPWL modeled surfaces. (a) Experimentally measured HEMT characteristic ($P_o = 1$ mW, $V_{gso} = 0$ V, $V_{dso} = 2$ V) (squares) and SPWL modeled surface (circles). (b) Derivative with respect to v_{ds} at the same point (squares) and SPWL estimate of the derivative with respect to v_{ds} (circles). (c) Derivative with respect to v_{gs} at the same point (squares) and SPWL estimate of the derivative with respect to v_{gs} (circles).

TABLE II
COMPUTATIONAL BURDEN OF THE DIFFERENT MODELS EXPRESSED BY THE TRAINING TIME

Model	Training Time (minutes)
MLP(5)	240
CPWL(5)	8
SPWL(5)	12
MLP(9)	540
CPWL(10)	14
SPWL(10)	22
MLP(11)	660
CPWL(12)	17
SPWL(12)	26
GRBF(8)	180

MLP(N): Multilayer Perceptron with N neurons in the hidden layer.

GRBF(N): Generalized Radial Basis Function with N neurons.

CPWL(N): Canonical Piecewise Linear Model with N boundaries in the input space.

SPWL(N): Smoothed Piecewise Linear model with N boundaries in the input space.

lows one to get a smooth and derivable approximation of the nonlinear I/V characteristic function with a low number of parameters and with a low computational burden when compared with other alternatives applied to this problem. Moreover, it has been shown that the smoothness (defined as a functional involving second derivatives) varies monotonically with a single smoothing parameter of the function. In this way, the use of the SPWL model inherently provides some kind of regularization of the modeling problem. The smoothing parameter can be selected to fit a derivative of the I/V characteristic, while the rest of the parameters of the model are selected to fit the function; this opens a way to obtain large-signal models with the capability of reproducing the small-signal intermodulation distortion behavior. However, more work in this line is needed to achieve this goal.

The proposed SPWL model has been applied to model a HEMT transistor under optical illumination using real measurements. Its comparison with some neural networks-based alternatives shows clear advantages both in terms of performance and computational burden.

APPENDIX A
TRAINING OF THE SPWL MODEL

Let us consider that we want to approximate a mapping (g: $R^M \rightarrow R$) using a set of N input–output samples $(\boldsymbol{x}_l, y_l)$, $l = 1, \ldots, N$, with $\boldsymbol{x_l} = (x_{1,l}, x_{2,l}, \ldots, x_{M,l})^T$. One coefficient from each boundary can be eliminated by rewriting

$\langle \boldsymbol{\alpha_i}, \boldsymbol{x} \rangle - \beta_i$, as

$$b_i(\boldsymbol{x}) = m_{i,1}x_1 + m_{i,2}x_2 + \cdots + m_{i,M-1}x_{M-1} - x_M + t_i \tag{21}$$

where $b_i(\boldsymbol{x})$ denotes the ith boundary evaluated at $\boldsymbol{x}$. 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

$$f(\boldsymbol{x}) = a + \boldsymbol{b}^T\boldsymbol{x} + \sum_{i=1}^{\theta} c_i \operatorname{lch}(b_i(\boldsymbol{x}), \gamma). \tag{22}$$

The model parameters can be grouped into two vectors: $\boldsymbol{z}_p$ and $\boldsymbol{z}_r$, which correspond to the linear coefficients and the boundaries of the input space, respectively

$$\boldsymbol{z}_p = (a, b_1, \ldots, b_M, c_1, \ldots c_\theta)^T \tag{23}$$

$$\boldsymbol{z}_r = (m_{1,1}, \ldots, m_{1,M-1}, \ldots, m_{\theta,1}, \ldots m_{\theta,M-1}, t_1, \ldots t_\theta)^T. \tag{24}$$

The error function to be minimized is given by

$$E(\boldsymbol{z}_p, \boldsymbol{z}_r) = \sum_{l=1}^{N} \left(y_l - \left(a + \boldsymbol{b}^T\boldsymbol{x}_l + \sum_{i=1}^{\theta} c_i \operatorname{lch}(b_i(\boldsymbol{x}_l), \gamma) \right) \right)^2. \tag{25}$$

For a set of fixed boundaries $\boldsymbol{z}_r$, the optimal $\boldsymbol{z}_p$ parameters are given by

$$\boldsymbol{z}_p = (\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\boldsymbol{y} \tag{26}$$

where $\boldsymbol{y} = (y_1, \ldots, y_N)^T$, $\boldsymbol{A}$ is the following $M + \theta + 1 \times N$ matrix:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{U} \\ \boldsymbol{V} \end{bmatrix} \tag{27}$$

where

$\boldsymbol{1}$ row of N ones;

$\boldsymbol{U}$ $M \times N$ matrix with elements, $u_{i,j} = x_{i,j}$;

$\boldsymbol{V}$ $\theta \times N$ matrix with elements $v_{i,j} = \operatorname{lch}(b_i(\boldsymbol{x}_j), \gamma)$.

The searching direction to modify $\boldsymbol{z}_r$ is evaluated by a second-order method

$$\boldsymbol{s} = -\boldsymbol{Y}^{-1}\boldsymbol{g} \tag{28}$$

where $\boldsymbol{g}$ is the gradient and $\boldsymbol{Y}$ the Hessian of (25). They are evaluated by

$$\boldsymbol{g} = 2\boldsymbol{K}\boldsymbol{G}\boldsymbol{e} \tag{29}$$

$$\boldsymbol{Y} = 2\boldsymbol{K}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{K} + 2\boldsymbol{K}\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{z}_r}\boldsymbol{e} \tag{30}$$

where $\boldsymbol{e} = (e_1, \ldots, e_N)^T$ is the vector of errors, and $\boldsymbol{K}$ and $\boldsymbol{G}$ are

$$\boldsymbol{K} = \operatorname{diag}\Big(\underbrace{c_1, \ldots, c_1}_{M-1 \text{ terms}} \underbrace{c_2, \ldots, c_2}_{M-1 \text{ terms}} \cdots \underbrace{c_\theta, \ldots, c_\theta}_{M-1 \text{ terms}}, c_1, c_2, \ldots, c_\theta \Big) \tag{31}$$

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{G}^1 \\ \vdots \\ \boldsymbol{G}^\theta \\ \boldsymbol{P} \end{bmatrix} \tag{32}$$

where $\boldsymbol{G}^k$ are $M - 1 \times N$ matrices with $g^k_{i,j} = x_{i,j} \tanh(\gamma b_k(\boldsymbol{x}_j))$, and $\boldsymbol{P}$ is an $\theta \times N$ matrix with elements $p_{i,j} = \tanh(\gamma b_i(\boldsymbol{x}_j))$.

The second term of the Hessian in (30) involves the second derivative of the SPWL model, $\operatorname{sech}^2(b_i(\boldsymbol{x}_l))$, 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 this term from the Hessian; that is $\boldsymbol{Y} = 2\boldsymbol{K}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{K}$.

The new boundaries are estimated as

$$\boldsymbol{z}_r = \boldsymbol{z}_r + \alpha \boldsymbol{s} \tag{33}$$

where $\alpha = \arg\min(E(\boldsymbol{z}_p, \boldsymbol{z}_r + \alpha \boldsymbol{s}))$.

REFERENCES

[1] 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 Technol.*, vol. 33, pp. 1383–1394, 1985.

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

[3] S. A. Maas and A. Crosmun, "Modeling the gate I/V characteristic of a GaAs MESFET for Volterra-series analysis," *IEEE Trans. Microwave Theory Technol.*, vol. 37, pp. 1134–1136, 1989.

[4] D. E. 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 European Microwave Conf.*, Stuttgart, Germany, 1991, pp. 923–927.

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

[6] J. Rousset, Y. Harkouss, and J. M. Campovecchio, "An accurate neural network model of FET for intermodulation and power analysis," in *Proc. 26th European Microwave Conf.*, Prague, Czechoslovakia, 1996.

[7] K. Shirakawa, M. Shimiz, N. Okubo, and Y. Daido, "A large signal characterization of a HEMT using a multilayered neural network," *IEEE Trans. Microwave Theory Technol.*, vol. 45, pp. 1630–1633, 1997.

[8] A. H. Zaabab, Q. J. Zhang, and M. Nakhla, "A neural network modeling approach to circuit optimization and statistical design," *IEEE Trans. Microwave Theory Technol.*, vol. 43, pp. 1349–1358, 1995.

[9] I. Santamaría, M. Lázaro, C. J. Pantaleón, J. A. García, A. Tazón, and A. Mediavilla, "A nonlinear MESFET model for intermodulation analysis using a generalized radial basis function network," *Neurocomputing*, vol. 25, pp. 1–18, 1999.

[10] J. A. García, A. Tazón, A. Mediavilla, I. Santamaría, M. Lázaro, C. J. Pantaleón, and J. C. Pedro, "Modeling MESFETs and HEMTs for intermodulation distortion behavior using a generalized radial basis function network," *Int. J. RF Microwave CAE*, vol. 9, pp. 261–276, 1999.

[11] L. O. Chua and A. C. Deng, "Canonical piecewise-linear modeling," *IEEE Trans. Circuits Syst.*, vol. CAS 33, pp. 511–525, 1986.
[12] ——, "Canonical piecewise-linear representation," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 101–111, 1988.
[13] J. N. Lin and R. Unbehauen, "Canonical representation: From piecewise-linear function to piecewise-smooth functions," *IEEE Trans. Circuits Syst.*, vol. 40, pp. 461–468, 1993.
[14] D. M. W. Leenaerts and W. M. G. van Bokhoven, *Piecewise Linear Modeling and Analysis*. Boston, MA: Kluwer, 1998.
[15] D. W. Bacon and D. G. Watts, "Estimating the transition between two intersecting straight lines," *Biometrika*, vol. 58, pp. 525–534, 1971.
[16] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. New York: Wiley, 1989.
[17] D. A. Griffiths and A. J. Miller, "Hyperbolic regression—A model based on two phase piecewise-linear regression with a smooth transition between regimes," *Commun. Stat.*, vol. 2, pp. 561–569, 1973.
[18] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, pp. 1481–1497, 1990.
[19] E. F. Calandra and G. Sirna, "CAD-oriented modeling of the optically-controlled GaAs MESFET," in *Proc. GAAS'94 Symp.*, Torino, 1994, pp. 401–404.
[20] A. Madjar, P. R. Herczfeld, and A. Paolella, "Analytical model for optically generated currents in GaAs MESFETs," *IEEE Trans. Microwave Theory Technol.*, vol. 40, pp. 1681–1691, 1992.
[21] C. Navarro *et al.*, "Large signal dynamic properties of GaAs MESFET/HEMT devices under optical illumination," in *Proc. GAAS'98 Symp.*, Amsterdam, 1998, pp. 350–353.

Marcelino Lázaro was born in Carriazo, Spain, in 1972. He received the telecommunication engineer degree from the Universidad de Cantabria, Spain, in 1996. In 1996, he joined the Departamento de Ingeniería de Comunicaciones at the Universidad de Cantabria, where he is currently pursuing the doctor degree. His research interest includes digital signal processing, nonlinear modeling, and neural networks.

Ignacio Santamaría was born in Vitoria, Spain, in 1967. He received the telecommunication engineer degree and the doctor degree from the Universidad Politécnica de Madrid (UPM), Spain, in 1991 and 1995, respectively. In 1992, he joined the Departamento de Ingeniería de Comunicaciones at the Universidad de Cantabria, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, nonlinear systems, and neural networks.

Carlos Pantaleón was born in Badajoz, Spain, in 1966. He received the telecommunication engineer degree and the doctor degree from the Universidad Politécnica de Madrid (UPM), Spain, in 1990 and 1994, respectively. In 1990, he joined the Departamento de Ingeniería de Comunicaciones at the Universidad de Cantabria, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, nonlinear systems, and neural networks.

Angel Mediavilla Sánchez (M'92) was born in Santander, Spain, in 1955. He graduated (with honors) in 1978 and received the Doctor of physics degree in 1984, both from the University of Cantabria, Santander, Spain. From 1980 to 1983, he was Ingénieur Stagiére at Thomson-CSF, Corbeville, France. Now, he is Professor at the Department of Electronics of the University of Cantabria.

He has a wide experience in analysis and optimization of nonlinear microwave active devices and circuits in both hybrid and monolithic technologies. He has participated in Spanish and European projects in nonlinear modeling (Esprit project 6050 MANPOWER) and microwave and millimeter wave communication circuits and systems (Spanish Project PlanSAT, European Project CABSINET, etc).

His current research fields are on active microwave circuits, mainly in the area of nonlinear modeling of GaAs devices and their applications in large-signal computer design.

Antonio Tazón Puente (M'92) was born in Santander, Spain, in 1951. He graduated in 1978 and received the Doctor of physics degree in 1987, both from the University of Cantabria, Santander, Spain. From 1991 to 1995, he was Professor at the Department of Electronics University of Cantabria) and since 1996 he is Professor at the Department of Communication Engineering also of the University of Cantabria, Spain. In 1985, from March to October, in 1986, from April to July, he carried out stages at the IRCOM Department (University of Limoges, France) working in nonlinear modeling and load-pull techniques.

He has participated in Spanish and European projects in the nonlinear modeling (Esprit project 6050 MANPOWER) and microwave and millimeter wave communication circuits and systems (Spanish Project PlanSAT, European Project CABSINET, etc).

He has carried out research on analysis and optimization of nonlinear microwave active devices and circuits in both hybrid and monolithic technologies. Currently he main research interests is the active microwave circuits, mainly in the area of linear and large-signal modeling and small signal intermodulation of GaAs and Si–Ge devices and their applications in nonlinear computer design.

Tomás Fernández was born in Torrelavega, Spain, in 1966. He received the Ph.D. degree in physics from the University of Cantabria in 1996. Since 1998, he has been an Assistant Lecturer in the Department of Communication Engineering at the University of Cantabria. His research interests include large signal, nonlinear modeling of III–V compound transistor as well as SiGe HBT devices.