DATA-DRIVEN MODELING OF NON-LINEAR MICROWAVE DEVICES

Bernd Schoner and Neil A. Gershenfeld
Physics and Media Group
MIT Media Laboratory
Cambridge, MA 02139
{schoner,neilg}@media.mit.edu

Abstract

An inference-based method for modeling complex non-linear systems is presented that integrates current approaches to modeling of microwave devices within a generalized non-linear framework. Familiar techniques for characterizing and emulating linear and non-linear systems are embedded in an automated weighting mechanism, so that globally complex behavior is approximated by a number of simple local models. The framework handles non-Gaussianity, non-stationarity and discontinuity. It offers meaningful parameters, such as linear filter coefficients or Volterra expansion coefficients, as well as detailed error estimates.

1 Introduction

Microwave device technology has evolved into application domains where its emulation and characterization with linear techniques are no longer satisfactory. Although non-linear effects appear in many systems to such an extent that failure or success of ambitious design goals depend on them, the measurement technology as well as the data representation in state of the art analyzer systems still relies on the assumption of linearity. As a result, most of the interesting and useful non-linear behavior of microwave devices is either missed or neglected.

Non-linearity is difficult to model as well as to characterize; while physical models [3] require impractical amounts of computation, reliable electrical models are device-specific and tedious to design [6]. Motivated by successful inference approaches to system characterization in other engineering domains, we present an approximation architecture that characterizes a device based on analysis of empirical data collected from the device under test (DUT). The framework is based on simple local models that share the input data space, weighted by global Gaussian kernels. Although the model's local structure is based on conventional well-understood modeling practice and theory, it is able to predict globally complex and non-linear behavior.

While section 2 of this work presents the framework as a general function approximation framework, section 3 introduces two local architectures that are specifically suitable to modeling microwave devices. The first architecture is concerned with a time domain approach. Linear systems theory has shown how linear autoregressive models can approximate a linear filter of arbitrary complexity given an arbitrary number of tabs in the autoregressive model (IR-Filter). Non-linear systems theory has shown how internal degrees of freedom of a physical system can be reconstructed from input and output observables of the system [8, 1]. This reconstructed state space can be used to predict arbitrarily non-linear systems. In between these two extreme cases there are many systems that behave linearly at a local scale but show non-linear behavior when driven over a wide range of input signals. As an example of such a system we introduce a data set of simulated amplitude modulated input and output signals from an Ebers-Moll transistor model (fig.5) [11]. This circuit can be efficiently described and predicted by linear models that are embedded in the non-linear function approximation framework.

References [9] and [10] introduced the non-linear S-parameter equivalent for weakly non-linear multi-port devices. As described by the Volterra theory, non-linear interaction between frequency components can be modeled by polyno-

Figure 1:

mials beyond the first order linear terms. Such model is able to characterize harmonic responses at the output-port due to non-linear coupling between harmonic components at the input-ports, up to an arbitrary order of approximation. We show how the Volterra approach to device modeling is integrated into our global framework. Although Volterra approximations perform well on the local scale, their accuracy degrades over a larger dynamic range and a wider frequency span. Once again our algorithm overcomes this problem, as it optimally allocates a series of local models describing different input-output behavior.

2 Global Architecture

To start, we take a set of experimental measurements by driving a device with signals $\mathbf{v}_n$ at the input port and measuring the output signals $\mathbf{w}_n$ at the output-port. $\mathbf{v}_n$ is drawn from $p_v(\mathbf{v})$, where $p_v(\mathbf{v})$ describes the unconditioned probability of $\mathbf{v}$ given a specific application. Let $\mathbf{x}$ and $\mathbf{y}$ be any quantities of interest in the time or frequency domain derived from $\mathbf{v}$ and $\mathbf{w}$, such as complex coefficients describing the incident, transmitted, and reflected energies at the device (fig. 1).

Given the set of measurements $\{\mathbf{y}_n, \mathbf{x}_n\}_{n=1}^N$, we infer the joint probability density $p(\mathbf{y}, \mathbf{x})$, which lets us derive conditional quantities such as the expected value of $\mathbf{y}$ given $\mathbf{x}$, $\langle \mathbf{y}|\mathbf{x} \rangle$, and the expected covariance matrix of $\mathbf{y}$ given $\mathbf{x}$, $\langle \mathbf{C}_y|\mathbf{x} \rangle$. The value $\langle \mathbf{y}|\mathbf{x} \rangle$ serves as prediction of the target value $\mathbf{y}$ and $\langle \mathbf{C}_y|\mathbf{x} \rangle$ serves as its error estimate [4].

The joint density $p(\mathbf{x}, \mathbf{y})$ is expanded in clusters labeled c_m, each of which contains an input domain of influence, a local model, and an output distribution:

$$p(\mathbf{y}, \mathbf{x}) = \sum_{m=1}^{M} p(\mathbf{y}, \mathbf{x}, c_m) \tag{1}$$

$$= \sum_{m=1}^{M} p(\mathbf{y}|\mathbf{x}, c_m)\ p(\mathbf{x}|c_m)\ p(c_m)$$

The kernel probability functions $p(\mathbf{y}|\mathbf{x}, c_m)$ and $p(\mathbf{x}|c_m)$ are taken to be Gaussian so that $p(\mathbf{x}|c_m) = \mathcal{N}(\mu_m, \mathbf{C}_m)$ and $p(\mathbf{y}|\mathbf{x}, c_m) = \mathcal{N}(\mathbf{f}(\mathbf{x}, \beta_m), \mathbf{C}_{y,m})$, where $\mathcal{N}(\mu, \mathbf{C})$ stands for the multi-dimensional Gaussian distribution with mean vector μ and covariance matrix $\mathbf{C}$. The function $f(\mathbf{x}, \beta_m)$ with unknown parameters β_m is taken to be a linear coefficient model of the form

$$\mathbf{y} = \sum_{m=1}^{M} \beta_m \mathbf{f}_m(\mathbf{x}_n) \tag{2}$$

Given this density estimate we infer a conditional forecast

$$\langle \mathbf{y}|\mathbf{x} \rangle = \frac{\sum_{m=1}^{M} \mathbf{f}(\mathbf{x}, \beta_m)\ p(\mathbf{x}|c_m)\ p(c_m)}{\sum_{m=1}^{M} p(\mathbf{x}|c_m)\ p(c_m)} \tag{3}$$

as well as a conditional error forecast,

$$\langle \mathbf{C}_y|\mathbf{x} \rangle = \tag{4}$$
$$\frac{\sum_{m=1}^{M} [\mathbf{C}_{m,y} + \mathbf{f}(\mathbf{x}, \beta_m)\mathbf{f}(\mathbf{x}, \beta_m)^T]\ p(\mathbf{x}|c_m)\ p(c_m)}{\sum_{m=1}^{M} p(\mathbf{x}|c_m)\ p(c_m)}$$
$$-\langle \mathbf{y}|\mathbf{x} \rangle^2$$

3 Local Models

3.1 Time Domain Approach

So far the local models $f(\mathbf{x})$ have only be constrained to be of form (2). In this section we specify a local architecture that allows one to model weakly non-linear multi-port networks in the time domain.

If the output time-series at time t and at a given operating point is a linear function of the input signals at time t, the output model is of the form

$$y(t) = \beta_0 + \sum_{d=1}^{D} \beta_d \cdot x_d(t) \tag{5}$$

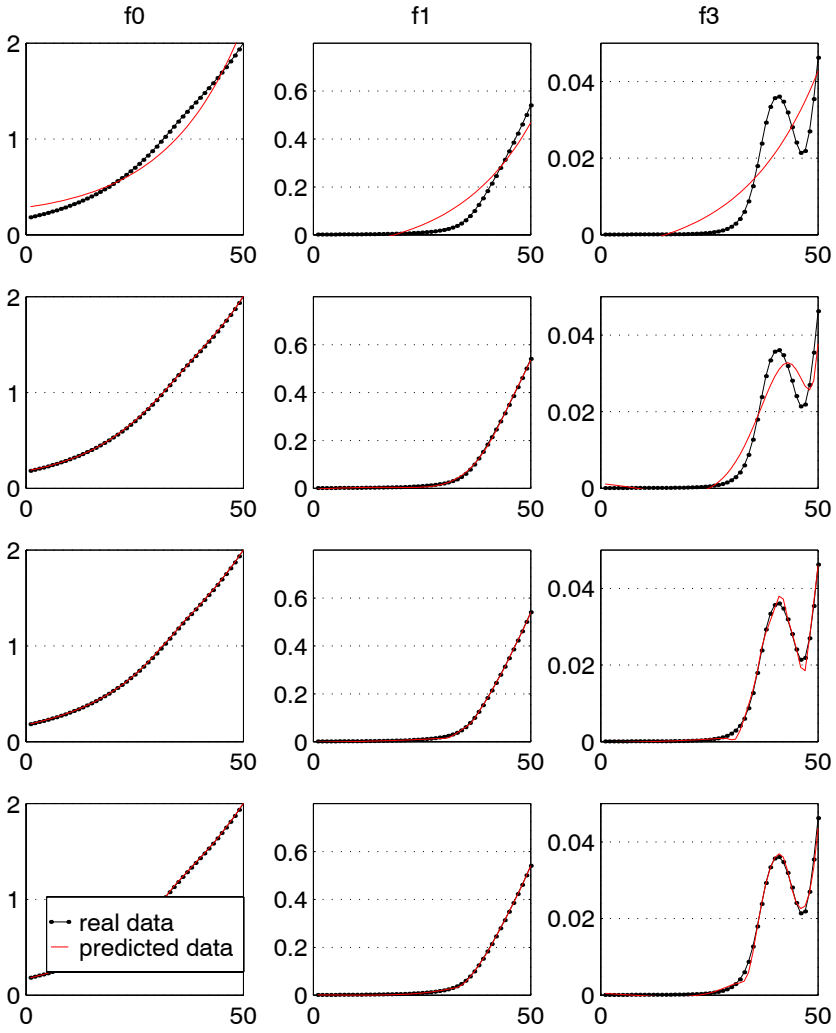

Figure 2: Approximation of 3 (out of 4) frequency components. The rows represent the approximation results of (a) a linear model, (b) a 5th-order-polynomial, (c) 5-cluster/local-linear model, (d) 3-clusters/local-quadratic model.

If the system has some sense of memory we may use lagged values of the input time series to augment the state vector. In this approach of 'weak' state space reconstruction the model becomes

$$y(t) = \beta_0 + \sum_{d=1}^{D} \sum_{k=1}^{K} \beta_{d,k} \cdot x_d(t - k \cdot \tau) \qquad (6)$$

where k is the number of lags and τ is the time lag. This approach avoids the stability problems of the embedding method, but retains the notion of state augmentation and allows for present output to be conditioned on past system states [1].

3.2 S-Parameter Approach

In this section we specify a local architecture to represent and predict RF multi-port networks in the frequency domain. A linear multi-port network is completely specified by its scattering and reflection parameters $S(\omega)_{i,j}$, where i refers to an input and j to an output signal, and the device load can be described by linear superposition of input signal components. However, transfer and reflection coefficients of non-linear devices can only be specified as functions of all input frequency components.

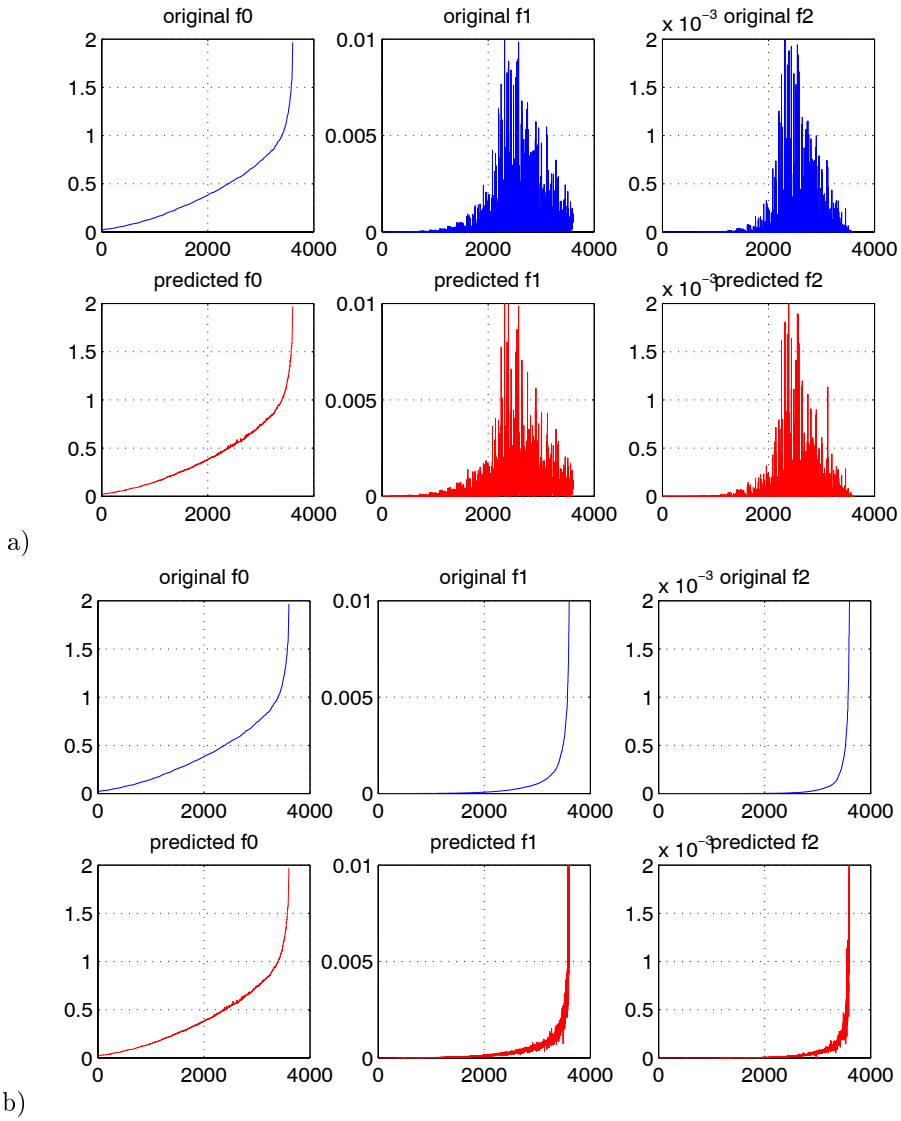

Figure 3: Approximated data from a simulated amplifier $(b_2(1), b_2(2), b_2(3))$. The x-axis gives the number of a test run, the y-axis indicates the energy in the component. Original and predicted output data a) sorted by the energy of the fundamental, b) sorted individually by the energy of the particular component. The noise in the predicted $f1$ and $f2$ plot indicates a slight prediction error.

Starting from fig.1 we need to identify the effect of combinations of input signals a_i on the output signals b_j. Restricting input and output to harmonic signal components, we denote the input component associated with the fundamental f_1, and the harmonics $f_2, f_3...$ by $a_i(1), a_i(2), ..$ respectively. We designate the corresponding output components by $b_i(1), b_i(2), ...$

The S-parameter approach for linear microwave devices is extended into non-linear transmission kernels H_n using the Volterra theory [10, 7]. $H_{n,ji_1i_2..i_n(f_1,f_2,..f_n)}$ describes the n-th order effect of frequency components f_k at the input port i_k on the frequency components $f_1 + f_2 + .. + f_n$ at output port j, where conjugate complex components are denoted by a negative f_i [10]. While the pure Volterra approach requires a detailed expansion of the Volterra series, we avoid this step by allowing any polynomial interaction between real and complex parts of signal components up to the

desired order O. The increase in the number of terms is compensated by a gain in symmetry that facilitates the parameter search.

Thus the local model is defined as

$$\mathbf{y} = \sum_{e_1+e_2+..+e_D<=O} \beta \cdot x_1^{e1} \cdot x_2^{e_2} \cdot ... \cdot x_D^{e_d} \quad (7)$$

and the order of the local model is traded off with the complexity of the global architecture (fig.2). While the polynomial expansion is very efficient within a limited input domain of interest, wide ranges of frequency and amplitude are best captured by an efficient split of the application domain into sub-domains of influence.

4 Parameter Estimation

The model parameters are found from a variant of the Expectation-Maximization (EM) algorithm, which computes the most likely cluster parameters by iterating between an expectation step and a maximization step [2, 5]. Conventional EM updates are used to estimate the unconditioned cluster probabilities $p(c_m)$, cluster locations μ_m and covariances $\mathbf{C}_m$. Pseudoinverses of the cluster weighted covariance matrices are used to update the local model parameters β_m.

E-step: Given a starting set of parameters, we find the probability of a cluster given the data:

$$\begin{aligned} p(c_m|\mathbf{y},\mathbf{x}) &= \frac{p(\mathbf{y},\mathbf{x}|c_m)\ p(c_m)}{p(y,\mathbf{x})} \quad (8) \\ &= \frac{p(\mathbf{y},\mathbf{x}|c_m)\ p(c_m)}{\sum_{l=1}^{M} p(\mathbf{y},\mathbf{x}|c_l)\ p(c_l)} \end{aligned}$$

where the sum over clusters in the denominator lets clusters interact and specialize in data they best explain.

M-step: Now we assume the data distribution correct and maximize the likelihood function changing the cluster parameters, starting with the cluster weights:

$$\begin{aligned} p(c_m) &= \int p(c_m|\mathbf{y},\mathbf{x})\ p(\mathbf{y},\mathbf{x})\ d\mathbf{y}\ d\mathbf{x} \quad (9) \\ &\approx \frac{1}{N}\sum_{n=1}^{N} p(c_m|\mathbf{y}_n,\mathbf{x}_n) \\ &\approx \frac{1}{N\ p(c_m)}\sum_{n=1}^{N} \theta(\mathbf{x}_n)\ p(c_m|\mathbf{y}_n,\mathbf{x}_n) \end{aligned}$$

Given $p(c_m)$ we define the cluster-weighted expectation of any function $\theta(\mathbf{x})$ as

$$\langle\theta(\mathbf{x})\rangle_m \equiv \int \theta(\mathbf{x})\ p(\mathbf{x}|c_m)\ d\mathbf{x} \quad (10)$$

which lets us update the cluster means and the cluster weighted covariance matrices :

$$\begin{aligned} \mu_m &= \langle\mathbf{x}\rangle_m \quad (11) \\ [\mathbf{C}_m]_{ij} &= \langle(x_i-\mu_i)(x_j-\mu_j)\rangle_m \end{aligned}$$

The derivation of the maximum likelihood solution for the model parameters yields

$$\beta_m = \mathbf{B}_m^{-1} \cdot \mathbf{A}_m \quad (12)$$

with $[\mathbf{B}_m]_{i,j} = \langle f_i(\mathbf{x}) \cdot f_j(\mathbf{x})\rangle_m$ and $[\mathbf{A}_m]_{i,j} = \langle y_i \cdot f_j(\mathbf{x})\rangle_m$. Finally the output covariance matrices associated with each model are estimated:

$$\mathbf{C}_{m,yy} = \langle[\mathbf{y} - \mathbf{f}(\mathbf{x}, \underline{\beta}_m)] \cdot [\mathbf{y} - \mathbf{f}(\mathbf{x}, \underline{\beta}_m)]^T\rangle_m \quad (13)$$

We iterate between the E- and the M-step until the total likelihood of the data, as defined by the product of all data likelihoods (equ.1) does not increase further (fig. 4b).

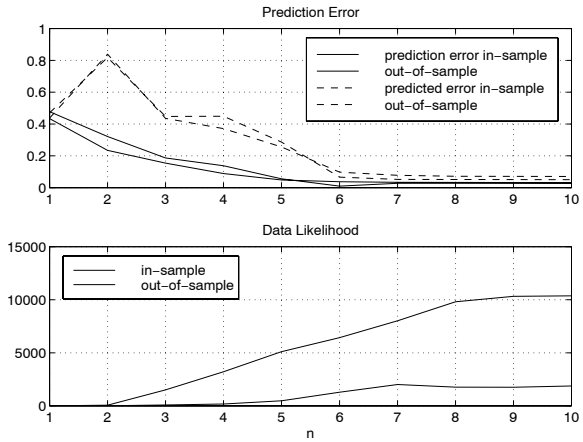

Figure 4: convergence of a 3-cluster/locally-quadratic Volterra coefficient model

.

5 Experimental Results

5.1 Time Domain Approach

Fig.5 illustrates the test device, simulated to generate data for the demonstration of the time domain approach: a bi-polar junction transistor is

embedded in a self-biased circuit functioning as a simple amplifier [11]. Different operating points of the device are selected by applying dc-potentials to V_{bb} and V_{cc} . An amplitude modulated signal is added to the dc-offset at V_{bb}. The modeling task consists in predicting the output-currents I_{BE} and I_{CE}, given the input voltages.

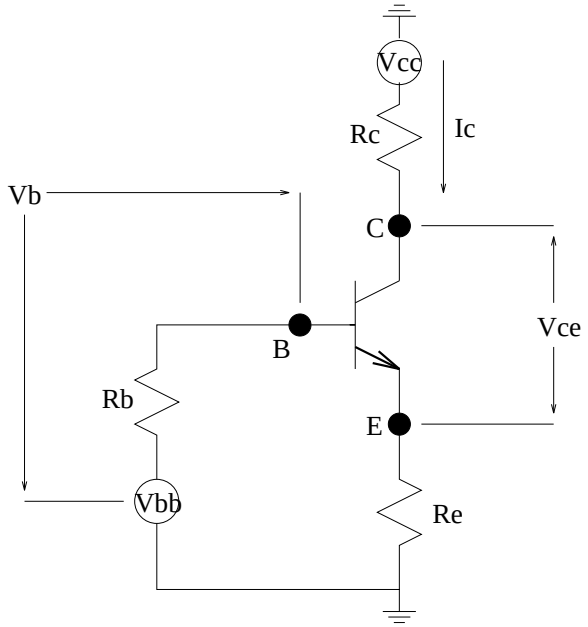

Figure 5: Test device for the time domain approach: a self-biasing transistor circuit. [11].

Fig.7a) shows how the data interpolates between bias points. For training a selection of bias points was used, while the test set was chosen from a larger number of bias points. The predicted out-of-sample test data is unrecognizibly similar to the true simulated output data and the relative RMS was smaller than 0.005 across the data set.

[11] uses radial basis functions to model input-output patterns for different bias points. Basis terms are fixed in specific input locations, so that bias points that haven't been observed in training can be computed by explicitly interpolating the basis functions of adjacent training points. Our result matches the error reported in [11]. Yet, in our system clusters automatically go where they are needed and interpolate appropriately. Hence our algorithm is not specific to the transistor example where bias points are chosen discreetly, but works for arbitrary devices. Also, the architecture remains general and applicable in different settings such as the one introduced in the next section.

5.2 S-Parameter Approach

Fig.(2) illustrates how local and global complexity contribute to the approximation result in the S-parameter approach. The data is taken from a device with a single input component at 1.5 GHz and 4 harmonic output components at 1.5, 3, 4.5 and 6 GHz. It is approximated by combinations of varying numbers of clusters and varying polynomial order. The purely linear approximation (2a) is unable to capture the data characteristics and the fifth order polynomial model (2b) still performs purely. The approximation with 5 linear models (2c) is doing well, while the approximation with only 3 second order polynomials is practically perfect (2c).

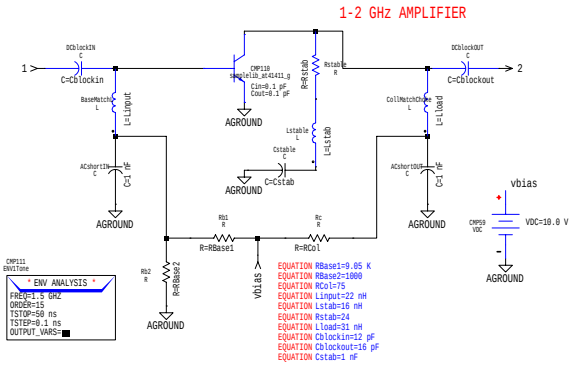

Figure 6: Test Device for S-parameter approach.

The model was further tested on data obtained from realistic simulations of an amplifier (fig.6) with two input-ports (input 1 and 2 at port 1) and a single output-port 2 [10]. There were one input component (1.5 GHz) at input-port 1 and three components (1.5, 3 and 4.5 GHz) at input-port 2, causing output components at 1.5, 3 and 4.5 GHz. Thus, the model predicts a 3-dimensional complex output vector, given a 4 dimensional complex input vector.

Multiple simulations were done with strongly varying boundary conditions. Fig.3 shows the approximation results from a particular simulation. The relative RMS error was at worst 0.001% for the fundamental and 0.5% for the second harmonic; on some of the test sets a significantly better performance was achieved. Local third order approxi-

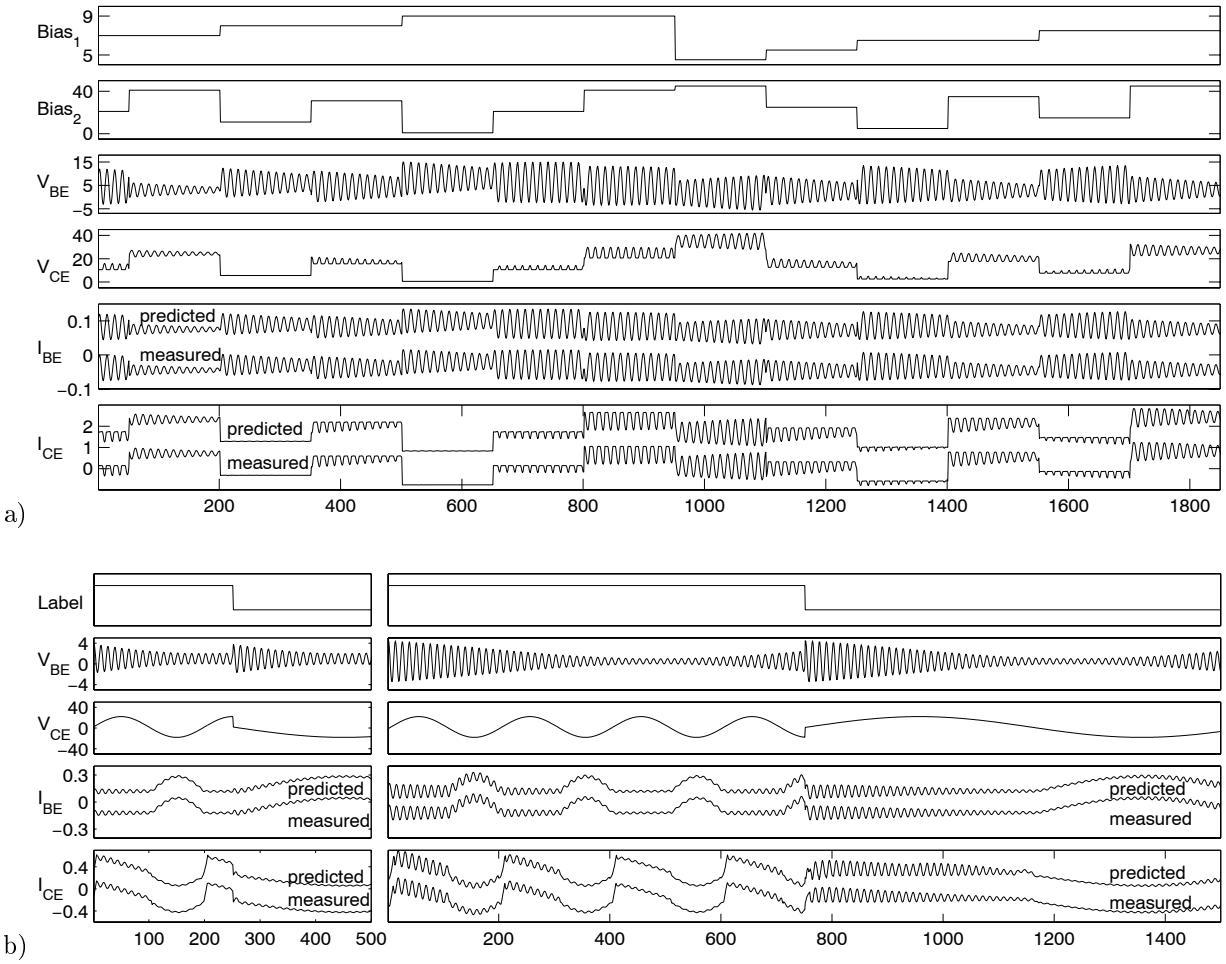

Figure 7: Time domain modeling. a) Amplitude modulated data from the ebers-moll transistor circuit: the two top rows represent the two bias points, the two middle rows the input voltages and the two bottom rows the predicted output currents, where the prediction is plotted above the measured (simulated) data to illustrate the similarity. b) data from the same circuit, but with stronger non-linear behavior: row 2 and 3 indicate the input voltages and the two bottom rows indicate the predicted output currents. The left column shows part of the training data, the right column shows test data.

mations performed best, which was expected given data ranging from the fundamental to the second harmonic. The difference between single and multiple cluster models should become more significant when absolute frequency will be varied and added to the feature vector.

6 Conclusions

It has been shown how the cluster-weighted architecture automatically allocates model parameters in the data space. It extends existing modeling techniques to complex systems that capture a device across any input range without losing performance in particular sub-domains. The experimental approximation of two particular amplifiers was shown to be arbitrarily accurate, given sufficient complexity of the model architecture.

We introduced two local architectures that extend the typical representation of current circuit simulation and test equipment. It was shown how our method predicts the output currents in the time domain, given different operating points of the device. This approach is appropriate for

systems that are insufficiently described by their steady state behavior. Furthermore it was demonstrated how the S-parameter characterization of linear devices can be extended to higher order polynomial approximations (Volterra series) and how this representation can be elegantly embedded in the cluster-weighted framework so that a wide range of driving signals is modeled in a single estimation step.

Our approach is yet to be tested on measurements of real world devices. Because our framework easily adapts to local and global data complexity, we expect it to outperform conventional non-linear techniques when applied to measured non-idealized and noisy data. Apart from the examples shown in this paper Cluster-Weighted Modeling can handle non-trivial systems that have discontinuous input-output characteristics or stochastic behavior.

Acknowledgments

The authors wish to thank Marc Vanden Bossche from HP Labs in Brussels, and Nicholas Tufillaro and Lee Barford from HP Labs in Palo Alto for providing the experimental data and for fruitful discussions. They also wish to thank Rich Fletcher for his useful comments. This work was made possible by the MIT Media Laboratory's Things That Think Consortium.

References

[1] Martin Casdagli. A dynamical systems approach to modeling input-output systems. In M. Casdagli and S. Eubank, editors, *Nonlinear Modeling and Forecasting*, Santa Fe Institute Studies in the Sciences of Complexity, pages 265–281, Redwood City, 1992. Addison-Wesley.

[2] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum Likelihood From Incomplete Data via the EM Algorithm. *J. R. Statist. Soc. B*, 39:1–38, 1977.

[3] F. Filicori, G. Ghione, and C.U. Naldi. Physics based electron device modeling and computer-aided mmic design. *IEEE Transactions on Microwave Theory and Techniques*, 40/7:1333–1352, 1992.

[4] N. Gershenfeld, B. Schoner, and E. Metois. Cluster-weighted modeling for time series analysis. *Nature*, 379:329–332, 1999.

[5] M.I. Jordan and R.A. Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6:181–214, 1994.

[6] B. Mallet-Guy, Z. Ouarch, M. Prigent, R. Quere, and J. Obregon. A distributed, measurement based, nonlinear model of fets for high frequencies applications. In *Microwave Symposium Digest of IEEE 1997 MTT-S International*, pages 869–872, New York, 1997. IEEE.

[7] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. John Wiley and Sons, Inc., New York, 1989.

[8] Floris Takens. Detecting strange attractors in turbulence. In D.A. Rand and L.S. Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381, New York, 1981. Springer-Verlag.

[9] F. Verbeyst and M. Vanden Bossche. Viomap, the s-parameter equivalent for weakly nonlinear rf and microwave devices. In *Microwave Symposium Digest of IEEE 1994 MTT-S International*, pages 1369–1372, New York, 1994. IEEE.

[10] F. Verbeyst and M. Vanden Bossche. The volterra input-output map of a high frequency amplifier as a practical alternative to load-pull measurements. In *Conference Proceedings TMTC'94*, pages 81–85, New York, 1994. IEEE.

[11] D.M. Walker, R. Brown, and N.B. Tufillaro. Constructing transpoartable behavioural models for nonlinear electronic devices. *Physics Letters A*. forthcoming.