

# A Behavioral Modeling Approach to Nonlinear Model-Order Reduction for RF/Microwave ICs and Systems

John Wood, *Senior Member, IEEE*, David E. Root, *Fellow, IEEE*, and Nicholas B. Tuffillaro

**Abstract**—This paper considers an approach to nonlinear model-order reduction for RF/microwave integrated circuits (ICs) from the perspective of “black-box” behavioral modeling. We present a systematic methodology for creating behavioral models using techniques developed from concepts in system identification, nonlinear dynamics, computational geometry, and information theory. Highly complex subsystems can be represented by relatively straightforward input–output relationships involving the observed and identified dynamical variables. Model order is thus significantly reduced compared with the device-level representation. We illustrate the technique by creating a cascadeable transportable model of a wide-band microwave IC amplifier that accurately predicts the dc, large-signal, harmonic and intermodulation distortion, and small-signal (*S*-parameter) behavior.

**Index Terms**—Estimation, identification, modeling, nonlinear systems, reduced-order systems.

## I. INTRODUCTION

ADVANCES IN nonlinear numerical simulation techniques have enabled the accurate design of new microwave and RF integrated circuits (ICs). Some key advances have been the development of modern harmonic-balance simulators and, more recently, of transient envelope simulators. Harmonic balance [1] allows the efficient simulation of large-signal steady-state circuits in the frequency domain, achieving great efficiency gains over traditional time-domain simulators like SPICE for steady-state problems with large numbers of frequency components. Such problems are commonly found in the design of microwave circuits. The recent availability of transient envelope simulators [2], [3] has allowed the efficient simulation of problems for which the typical spectra can be represented by a set of several discrete tones and time-dependent modulation around them. This is a common characteristic of modern digital communication circuits.

Modern microwave and wireless communications systems are too complex today to permit the complete simulation of the nonlinear behavior at the transistor level of description. This problem presents a significant productivity bottleneck for design engineers. A typical design and modeling hierarchy is depicted in Fig. 1. At the bottom is the device and at the top

is a complicated module or subsystem. A “top-down” design methodology propagates specifications down the hierarchy. Conversely, “bottom-up” verification is the process of validating overall system performance based on the performance of lower level components and their configuration. A solution to the design simulation bottleneck is to design at a higher level of abstraction at each level in the hierarchy. At the bridge between the transistor circuit and the multichip module or RF integrated circuit (RFIC) we use “behavioral models” to describe the nonlinear circuit blocks or ICs in the system. The behavioral models are simplified models of the essential nonlinear behavior of the complex sub-circuits; this simplification means that these models will execute more quickly, and use much less memory than if an entire complex subsystem was simulated at the transistor level. The critical need for nonlinear modeling techniques is a recent development driven by the increased size and complexity of ICs in the RF regime, as well as the adoption of more complex signal modulation techniques. The availability of such nonlinear modeling techniques will enable designers to make use of the advances in the simulation technology at higher levels of the design hierarchy.

In this paper, we describe a new and systematic time-domain methodology for generating nonlinear behavioral models that is based on techniques from nonlinear dynamics, system identification, and computational geometry. These behavioral models are a “black-box” approach to the problem of model-order reduction, as opposed to the traditional “white-box” approach, where detailed knowledge of the device physics or circuit configuration and operation is used to minimize the number of equations that describe the essential properties of the circuit or device. In this “black-box” approach, we are concerned only with describing the dynamical behavior of the circuit that is observed at its accessible terminals. The attraction of this measurement-based modeling approach is that a low-order model of a complex circuit or system can be derived, without prior knowledge of its internal circuitry or topology. Indeed, we advocate this approach even when such details are known, as it is based on the observable dynamics of the system, which are generally of much lower order than the internal states or dynamics. This approach is now beginning to be considered by the traditional model-order reduction community [4].

Our approach is similar to more conventional model-order reduction techniques in that both methods seek to identify and model the dynamics on a subspace of the phase space specified by the device level netlist. A dynamical model is created on this subspace that can also be implemented as a netlist. Our

Manuscript received December 12, 2003, revised May 28, 2004.

J. Wood and D. E. Root are with Worldwide Process and Technology Centers, Agilent Technologies Inc., Santa Rosa, CA 95403 USA (e-mail: john\_wood@agilent.com).

N. B. Tuffillaro is with Agilent Laboratories, Agilent Technologies Inc., Palo Alto, CA 94304 USA.

Digital Object Identifier 10.1109/TMTT.2004.834554

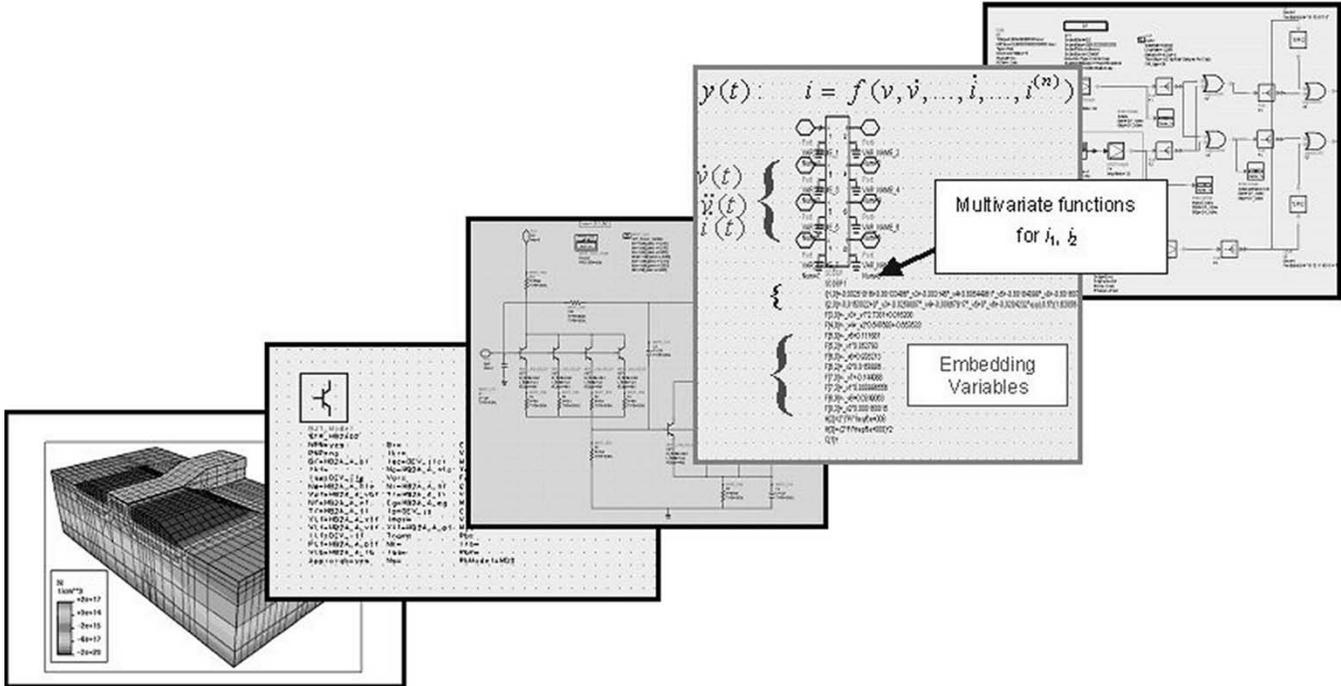


Fig. 1. Modeling hierarchy, beginning at the bottom with the device model described by the detailed semiconductor physics; this is abstracted to a circuit-level transistor model that describes the terminal behavior through equivalent-circuit or phenomenological equations. The transistor model is used to design circuits and ICs efficiently using circuit simulators, but it is too complex to use in the system-level simulation at the top of the hierarchy. To bridge the circuit and system simulation environments, we use a reduced-order behavioral model of the circuit.

method differs in that we work with the integrated solutions of this model provided by simulation, and not the netlist itself, i.e., we use simulations and not an analytic procedure to assist us in identifying the subspace and model used for order reduction. Both methods rely on the existence of a (small) subspace to realize efficiency gains.

We shall illustrate our methodology by creating and validating a model of a real microwave IC amplifier. The modeling procedures that we outline are very general: the test signal design, analysis, model generation, and simulator implementations are generic and can be applied to amplifiers, mixers, modulators, and other microwave components or subsystems. The resulting models are as follows:

- *transportable* [5]: in other words, usable in a range of system and simulation environments, and not restricted to a limited domain of applicability;
- *“cascadeable”* [6]—so that the cascade of two behavioral models performs faithfully with respect to the performance of the cascade of the respective transistor-level circuits.

Since cascading nonlinear components can create a wide variety of environments for the individual behavioral models, “cascadeability” implies a certain degree of transportability.

## II. METHODOLOGY

Fundamental to our approach is the notion that the behavioral model relates waveforms to waveforms, i.e., the output is not an instantaneous function only of the input signal. Rather, the output depends upon the shape of the input waveform or, equivalently, the output depends on the value of the input and past

values of the input, and even past values of the output itself: the concept of *memory*. We write this in functional notation according to (1). Here, we assume the voltage signal is the input and the current is the output

$$I(t) = F[V(t)]. \quad (1)$$

How we write the details of this functional depends on the nature of the system or component being modeled. Take, for example, a nonlinear resistor: the current is given by the instantaneous value of the applied voltage. The details, or the shape, of the voltage signal are unimportant, there are no dynamics in this system. In this case, the current–voltage relationship can be expressed as a simple algebraic function

$$I(t) = f(V(t)). \quad (2)$$

If we now place a capacitor in parallel with the resistor, the current flow depends on both the instantaneous voltage (across the resistor) and also the time derivative of the voltage (across the capacitor). The detail shape of the voltage signal is now important. The current is now expressed as a functional of the voltage signal, as given by (3) as follows:

$$I(t) = F[V(t)] = f(V(t), \dot{V}(t)). \quad (3)$$

This is an example of a *static* functional. The output depends upon the input signal only—a static relationship—but the output depends on the shape of the voltage signal, as expressed through the time derivative. The time derivative also describes a memory effect since a derivative can be approximated by a difference relationship or time delay.

Now consider a series connection of the resistor and capacitor (either or both of which may be nonlinear components). It can be shown fairly easily that the current through the  $RC$  network depends on the time derivative of the applied voltage and the time derivative of the current itself. We write this as an implicit relationship—a functional of the current and voltage signals. This more complicated class of models is given by the *feedback* or *dynamic* model of the form

$$F[V(t), I(t)] = 0 \quad (4)$$

where the solution can be expressed in terms of the output  $I(t)$

$$I(t) = f(\dot{V}(t), \dot{I}(t)). \quad (5)$$

Feedback models depend on an internal state of the system, and also embody the notion of memory.

The motivation for our approach to nonlinear systems identification goes under the rubric of nonlinear time series analysis (NL TSA) [7]. The suggestion to use this approach for describing input–output systems is due to Casdagli [8]. The key idea is to *embed* the measured or simulated stimulus and response variables in a higher dimensional space built not only from the measured data, but also transforms of the measured data, in our case, their time derivatives, which describe the local history of the signal. Due to a theorem of Takens, extended to the driven case by Stark [9], these embedded models can be faithful to the dynamics of the original system. In particular, deterministic prediction is possible from an embedded model that will mimic the dynamics of the actual system.

The models are formulated as implicit nonlinear ordinary time-differential equations, which are easily implemented in commercial microwave simulators, in the embedded variables

$$f(i(t), \dot{i}(t), \ddot{i}(t), \dots, v(t), \dot{v}(t), \ddot{v}(t), \dots) = 0. \quad (6)$$

The goal of the modeling process is thus to determine the significant embedding variables of the function  $f$  and then to find an efficient basis for the function approximation. Only the basic framework of our approach is presented here: the elaboration of the theory can be found in the references of this paper.

### III. EMBEDDING AS A MODEL-ORDER REDUCTION METHOD

The “black-box” model-order reduction procedure that we describe is similar to previous “white-box” techniques in that all model-order reduction methods seek significantly to reduce the number of variables (or states) used to predict the behavior of a device by some method of “projection.” The most common nonlinear model-order reduction techniques begin with a transistor level model (in the form of a netlist) and use a combination of functional approximations (e.g., Taylor series, Karhunen–Loeve expansions) and projection operators (e.g., Krylov subspace methods, Hankel norm approximations) to create a lower dimensional subspace that faithfully predicts the dynamics of interest [10]. For these methods to be successful, they must both identify a subspace (i.e., find a small set of

state variables) and create a projection operator that approximates the differential flow on this subspace (i.e., find a set of differential equations for the reduced set of state variables).

Our black-box method, in contrast, seeks to reconstruct both the subspace (the set of reduced state variables) and a suitable flow operator (a set of differential equations on this subspace) directly from a collection of measured or simulated data, i.e., we attempt to use the data itself to infer a model [11]. As previously mentioned, the insight for this approach comes from the Taken’s embedding theorem [9]. The big advantage of this approach is that it can be applied to systems for which the internal constitutive equations are unknown. This approach does, however, present some new challenges such as the design of a sufficiently rich data set that allows for the approximate reconstruction of the subspace and its model.

Roughly, previous nonlinear model-order reduction techniques are deductive in that they focus on the construction of a mathematical projection operator that starts from a netlist model. Our approach is inferential in that it focuses on approximating a model starting from a collection of data (experiments). White-box model-order reduction methods focus on the mathematical construction of a numerical projection operator assuming the prior model is correct. Black-box model-order reduction methods focus on inferring an approximate model assuming that the data is correct and sufficient to create a model useful for some well-defined set of excitations.

**Embedding** is both a simple procedure, and a profound insight into the behavior of the system [12]. We can write the dynamical description of a nonlinear system in terms of a set of nonlinear ordinary differential equations

$$\begin{aligned} \dot{\vec{x}}(t) &= \vec{f}(\vec{x}(t), \vec{v}(t)) \\ i(t) &= h(\vec{x}(t), \vec{v}(t)). \end{aligned} \quad (7)$$

Each of the  $x$ ’s in the vector  $\vec{x}(t)$  is a state variable of the system, the number of state variables describes the order of the system. The observable output variable  $i(t)$  is a function of the states of the system, and the external drive signal  $v(t)$ . This drive is also generally observable. The internal states  $\vec{x}(t)$  are not observable. If the state equations are known *a priori*, the value of the output  $i(t)$  can be determined for every time  $t$ . This solution describes a time-parameterized path or trajectory in the multi-dimensional space of the state variables known as the “phase space.” The observable output is a projection of this trajectory onto a single axis, the  $i$ -axis, and this describes the time evolution of the output value  $i(t)$ : plotted as a function of time, this is a time series.

Given a time series of some observable  $i(t)$ , a trajectory in a “model phase space” can be constructed using a process known as “embedding.” A common embedding procedure is to use delayed values of the observable output. This set of quantities constitutes an “embedding.” The actual system trajectory (of  $i(t)$ ) and the trajectory we create from this time-delay embedding will differ by no more than a smooth and differential change of coordinates—the transformation relating the actual and model trajectories is a *diffeomorphism*. In other words, the trajectory in the model space preserves the dynamics of the original system. What we have done is use an observable output to

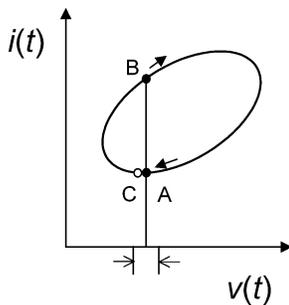


Fig. 2. Example  $i(t)$ – $v(t)$  relationship. The current is not a unique-valued function of the drive voltage.

retrieve explicitly the unseen internal degrees of freedom of the system—its internal state variables or their analogs. We can do this provided we take enough delays (or time derivatives, in the method described here) [7], [12].

We make use of a geometrical relationship to determine how many delays are needed. The Whitney embedding theorem states that an  $N$ -dimensional manifold can always be embedded in a Euclidean space of not more than  $2N + 1$  dimensions. For an example, consider a sheet of paper—a two-dimensional object. We can fold and attach the ends of the paper and get a Mobius strip, which requires three dimensions to describe it. If we then join the other ends, we get a Klein bottle, which requires four dimensions to describe it, but locally, on the piece of paper, the geometry is still two-dimensional—in the plane of the paper.

The “ $N$ -dimensional manifold” that we consider in behavioral modeling is the observable state space of the system. The embedding theorem provides an upper bound on the number of dimensions that we will need in our model state space. The utility of this approach is that, in practical cases, there is an orders-of-magnitude reduction in the number of variables (dimensions) required to describe the observed behavior of the system.

How many variables will we need in the embedding? The algorithm that we use for choosing which of the dynamical variables are used for the embedding is based on the technique of “false nearest neighbors” [13], which can be computed using algorithms from computational geometry. We have adapted a method described by Rhodes and Morari [14] for input–output systems. The algorithm uses the data itself to determine the optimal set of embedding variables, resulting in a compact and efficient model of vastly lower complexity than the original nonlinear system. The principle of this algorithm is illustrated in the following.

Consider the simple system comprising an observable output  $i(t)$  and a single drive signal  $v(t)$ , which yields the response shown in Fig. 2. Clearly the output is not a single-valued or unique function of the drive signal. For instance, the two points A and B on this curve share the same input value, but yield different outputs. Thus, if we sample the output  $i(t)$  in terms of the drive signal  $v(t)$ , then both samples A and B fall into the same “bin” from  $v$  to  $v + \Delta v$ : the points A and B are known as “false nearest neighbors” because they are close in the input space, but are from temporally disparate locations on the response curve. The point C, which is close to point A on the response curve—a true nearest neighbor—also falls into the same

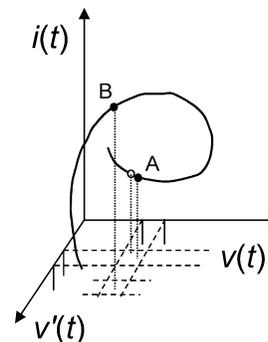


Fig. 3. Same example function now unfolded in a higher dimensional space of the embedding  $\{v(t), v'(t)\}$ . The current is now a single-valued functional of the embedding.

“bin.” The number of samples in each “bin” will, therefore, be large.

We can expand the number of embedding dimensions in this example by noticing that the direction of travel around the response curve means that the first time derivative of  $v(t)$  will be different at points A and B. We now plot the response  $i(t)$  as a function of the drive  $v(t)$  and the time derivative of the drive and, in this simple example, the response curve has unfolded into a single-valued path or trajectory (Fig. 3). Sampling  $i(t)$  in the new embedding space  $\{v(t), \dot{v}(t)\}$ , we see that the points A and B fall into separate “bins,” and the true nearest neighbor (point C) and point A still fall into the same “bin.” The number of counts in each “bin” has fallen.

This is the basic principle of the “false nearest neighbors” approach. We sample the observable output variable in the embedding space, initially assuming a simple one-dimensional model, and that most of the counts will be false nearest neighbors. As we add embedding variables, the state space is unfolded into higher and higher numbers of dimensions. At some point, the output response curve will unfold into a single-valued trajectory, and the only points in each bin will be true nearest neighbors. If the data is sampled appropriately, this will be a small number. We monitor the density of false nearest neighbors, and when this falls to a small value, this is the embedding dimension. This approach leads to fewer *ad hoc* assumptions, such as model order, compared with other recently published time-domain techniques [15].

The immediate differences between this approach and our application are that we are considering a driven system, which operates over a wide bandwidth. Clearly, a constant time-delay embedding is inadequate to cover the wide time scales (bandwidth) of the excitation signal used here for the amplifier. We use time derivatives of the inputs and outputs for embedding the data, yielding an expression of the general form

$$i(t) = f\left(v(t), \dot{v}(t), \dots, v^{(m)}(t), \dots, \dot{i}(t), \ddot{i}(t), \dots, i^{(n)}(t)\right). \quad (8)$$

Equation (8) is a feedback model of the type represented by (5). These behavioral models, therefore, can handle systems with memory. Equation (8) essentially defines an implicit nonlinear differential equation for the behavioral model: in other words, it describes the observable dynamical variables of the

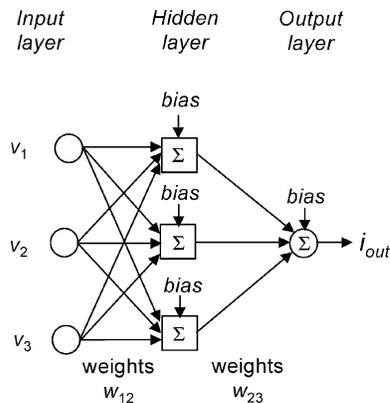


Fig. 4. ANN structure with a single hidden layer.

system. The remaining problem is to find a suitable approximation for the nonlinear function  $f$ .

#### IV. NONLINEAR FUNCTION FITTING USING ARTIFICIAL NEURAL NETWORKS (ANNs)

We now have a single-valued function relating the observable output and the embedding variables. The relationship between these variables is a nonlinear one, thus, we seek a multivariate nonlinear function fitting method. Multivariate polynomials [16], radial basis functions [17], and ANNs have been used [15], [18], [19]. ANNs are preferred due to their asymptotic properties and because they give very smooth results for approximating discrete measured and simulated data.

We use the basic structure shown in Fig. 4. The inputs are connected to the nonlinear processing units through a set of linear weights. The nonlinear units sum all their inputs, and produce an output when this sum is above a certain threshold, which can be adjusted by the bias. The transfer function for the processing units is a “sigmoid” function—hyperbolic tangent. The nonlinear behavior is captured in these functions. The outputs from all the processing units are summed through weights at the output.

A fundamental mathematical attraction for using ANNs is found in the “universal approximation theorem” [20], which states that, given enough neurons in the hidden layer, a neural network of the form shown can approximate any continuous bounded function to any accuracy that we care to specify.

Another feature of ANNs is “generalization”—the ability of a suitably trained network to correctly predict a response to a (set of) inputs that it has not seen before. In some cases, the network can be trained to fit the target data extremely well, but performs poorly on other data of a similar class—the network has “memorized” the target data and generalizes poorly. This is a symptom of overtraining.

Key to the design of the ANN for function approximation is the number of neurons in the hidden layer. Since the numbers of inputs and outputs are fixed—the former by the embedding procedure—the number of hidden layer neurons determines the number of weights that must be optimized during the training process to obtain the best function approximation. The values of the weights are obtained through “back-propagation,” a procedure where the network neuron outputs are used to

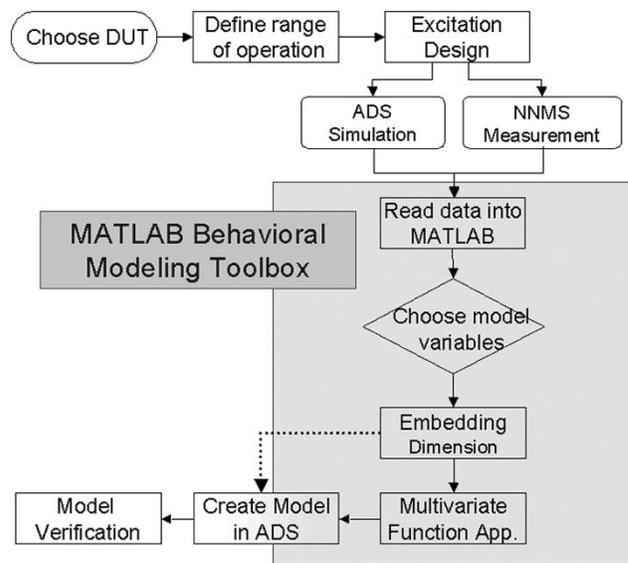


Fig. 5. NL TSA “black-box” modeling procedure. The shaded region identifies those functions that have been created using MATLAB. A suitable interface between MATLAB and the Agilent-EEsof ADS simulator has been devised to generate the SDD instance enabling the nonlinear model to be described in the simulator.

update the neuron input weight values through a minimization algorithm [21]: the Levenberg–Marquard nonlinear optimization algorithm is used. This minimization proceeds from the output of the network—comparing this with the target training value—to the input; hence, the terminology. The mean square error—the difference between the ANN output and the target value—is often used as a measure of the quality of the function approximation.

Techniques for improving the generalization of the ANN include regularization [21] and “early stopping” [21], [22]. Regularization uses other measures in the error term that is to be minimized, such as the sum of the squares of the weight values, to reduce the potential for overtraining of the network. Early stopping techniques use an additional set of data, a “validation” set, against which the ANN output is compared. If the error on the validation set begins to rise, even though the error on the training data set continues to fall, then this is an indication of a loss of generalization, and is an indicator to stop training the network. Cross-validation techniques [21], [22] use multiple data sets for training and validation to improve generalization using early stopping criteria. Bayesian techniques have been employed for weight selection to improve regularization [23], and these can be used to indicate the optimum network size. Early-stopping and Bayesian regularization methods work well on large ANNs, though this may not lead to parsimonious models and, hence, may lead to poor convergence in a simulator environment.

#### V. NL TSA BEHAVIORAL MODELING PROCEDURE

The modeling framework is outlined in Fig. 5. This illustrates the general flow of activities that we need to perform to extract a behavioral model from either measured or simulated data [24].

The first step is to define the excitation signals to the device-under-test (DUT) so that the device can be excited over its complete range of operation or over a limited range of interest.

The goal here is to choose (a set of) input signals such that all the observable nonlinear behavior of the device in a typical application is exercised. It is not necessary to design a signal that will excite all possible internal states in the device: indeed, some of these states will not be observable under the conditions of use and, in general, we do not know what the internal state space of the device is. Only the observable nonlinearities are of interest in the modeling process. The operating range of the DUT is generally specified in a datasheet. For example, the “power bandwidth” specification of the device or circuit can be used to define the range of powers and frequencies for the design of the (set of) excitation signals. The class of excitation signals that we have considered includes: 1) single-tone sinusoid; 2) two-tone or multitone sinusoids; and 3) analog or digitally modulated signals, whose carriers can be swept in power or frequency. These excitations can be produced in a simulation environment, or as practical signals using a signal generator with a large-signal network analyzer (LSNA) [25] to record all the frequency components present as harmonics and mixing products of the input signals resulting from the nonlinear behavior of the DUT [16], [18], [24]. A general theory of excitation design and optimal excitation design in this framework is still an open question. To date, we often apply an interactive procedure where we first identify an embedded phase space for model-order reduction and then examine the coverage (by examining the probability density of the excitation signals) on this phase space. We can then modify our excitation signals in an attempt to provide a relatively uniform coverage of the phase space.

The attraction of measurement-based modeling is that a low-order model of a complex component can be derived using this methodology without knowledge of the internal circuitry or topology of the component. Such an approach is not possible with more traditional modeling and model-order reduction techniques since these latter usually start with detailed circuit knowledge in the form of a netlist.

The measured or simulated stimulus and response data is imported into a prototype MATLAB-based Behavioral Modeling Toolbox where we perform the modeling procedures.

We typically use the DUT terminal voltages and currents, and their time derivatives up to second order or more, as the candidate variables from which to build the models. Typically, reported methods for building an embedded phase space from a nonlinear time series usually assume that there is a single input and single output, that the system can be described by a single characteristic timescale, and that new variables are created by delays [7]. Identifying which variables to use in the model is not a problem; a unique set of model variables is created from the delayed embeddings. Our problem differs in that we have multiple signals, and we have chosen to use time derivatives as candidate embedding variables to enable us to describe the wide frequency range covered by the DUT. From this candidate set of model variables, we need to select a subset from which to build a deterministic model. We start by using the “false nearest neighbors” method outlined earlier to identify a suitable set of embedding variables from these candidates. A “nearest neighbors” search algorithm from the TSToolbox<sup>1</sup> is used. Time-correlated

samples from the time series are excluded from the search for a given sample: only data points that are beyond a time interval that is found from the autocorrelation or auto mutual information of the signal are used in the search. All possible combinations of the voltages, currents, and their time derivatives are submitted to the search, and the false nearest neighbors returned as a percentage for each. For all candidate sets with a low percentage score, we fit a cubic polynomial to the nonlinear function, and estimate the residual error. The most promising candidate set(s) is(are) chosen, using compactness of the candidate set and minimum residual error as guides in this choice. This is an informal application of a minimum description length criterion [26].

Once the embedding has been identified, the nonlinear function approximation is carried out. As indicated earlier, we have tried polynomials, and radial basis function approximations, but typically use feed-forward ANNs. We use the MATLAB Neural Network Toolbox.<sup>2</sup> The embedding variables—voltages, currents, and their time derivatives—are inputs to the network. The network training is carried out using back-propagation and Levenberg–Marquard optimization: the training is stopped manually once the training error reaches a minimum and begins to plateau. While this often gives good results, the ANN may be less than optimal [27], finding a local rather than global minimum of the function approximation. More sophisticated training methods including regularization [21] and cross-validation [22] are under investigation.

The MATLAB neural-network structure is then parsed as a text file to be read into this nonlinear circuit simulator, i.e., Agilent-EEsof Advanced Design System (ADS), for verification of the model. A proprietary piece of software is used to convert this text file into an instance of a symbolically defined device (SDD) in ADS. An example can be seen in Fig. 1. The SDD calculates the time derivatives directly from its port voltages and currents at each time step during the simulation. A practical implementation also requires variable scaling, which is also implemented within the SDD. Validation of the model against measurement or simulation of the transistor-level circuit is carried out in ADS. Accuracy and speed of simulation are figures-of-merit for the behavioral model: the goal of a much shorter simulation time indicating that a reduced-order model compared with the full transistor-level circuit has been created.

It is helpful to point out the possible limitations of the method proposed. First, as noted earlier, the issues around good excitation design is still under active investigation. Furthermore, unlike model-order reduction techniques that are built on direct transformations of the netlist, we currently have no way to guarantee that the model will extrapolate well to predictions with excitations of a type the model is not trained on. However, especially for design problems for which the excitation signals are well defined, the proposed model-order reduction procedure can be very useful.

## VI. MICROWAVE AMPLIFIER IC MODEL

The DUT is a wide-band microwave IC amplifier, i.e., Agilent Technologies Inc.’s HMMC-5200 [28]. This is a dc–20-GHz

<sup>1</sup>[Online]. Available: <http://www.physik3.gwdg.de/tstool/index.html>

<sup>2</sup>[Online]. Available: <http://www.mathworks.com>

10-dB gain amplifier with internal feedback, designed to be used as a cascadable gain block in a variety of microwave circuit applications. It consists of eight HBTs configured as a compound modified Darlington feedback pair, operating in class A. The IC also contains biasing and feedback resistors, on-chip bypass capacitors, etc. The IC has numerous nodes and connections and, hence, has many internal degrees of freedom. Since one of the key assumptions of our methodology is that only a few of these internal states are important features of the observable signal, the resulting model should be quite compact; in other words, a significant reduction in the model order can be achieved.

The data for creating the behavioral model of this amplifier are generated from simulation. The excitation signal applied to this DUT is two offset tones at the amplifier input, and a tone at the output port, identical in frequency to one of the input tones. These signals were swept over the frequency range from 1.2 to 10.2 GHz, and from small signal to the amplifier's  $P - 1$  dB compression point: 0 dBm for each input tone. The signal power applied at the output port was identical to the input, i.e., approximately 10 dB below the output generated by the IC. The input tone separation used was 600 MHz, as employed in the LSNA instrument. We performed a harmonic-balance simulation using ADS; nine harmonics for each individual tone, and mixing [intermodulation (IM)] tones up to nine orders were considered. The amplifier requires a dc bias to be applied through the RF output port, via a load resistor and choke (dc feed component), and also requires dc blocks on the input and output ports. The voltages and currents, including dc contributions, were monitored at the amplifier's RF input and output ports. The harmonic-balance voltage and current data were converted into time series signals using the ADS fast Fourier transform (FFT) function.

From ADS, the data was exported to our prototype MATLAB Behavioral Modeling Toolbox, as shown in Fig. 5. As described above, we use as candidate embeddings the amplifier's terminal voltages and currents, and the time derivatives of these up to second order. Initially, we used only terminal-1 variables as candidates for the embedding variables for the terminal-1 current; this choice was made on the basis that the amplifier's  $S_{12}$  is quite small and, therefore, the effects of the output variables could be neglected. Subsequent studies have demonstrated that this hypothesis was false: these models behaved poorly due to the incorrect embedding choice, as terminal-2 variables were shown by the false nearest neighbors technique to be significant contributors to the terminal-1 current model. The amplifier model reported here uses all voltages and currents and time derivatives to second order as embeddings for both the terminal-1 and terminal-2 currents. These embeddings gave the minimum residual for a cubic polynomial fit, but may not necessarily be the most compact models.

The nonlinear functions for the terminal-1 and terminal-2 currents were individually approximated by ANN models. The neural networks used a single "hidden layer" of 40 neurons: this number is of similar order to that reported by Xu *et al.* [15] in the modeling of a narrow-band amplifier using ANNs. The neural networks were trained for approximately 15–25 cycles before stopping manually: at this point, the mean square error had fallen to a low value, and further reductions were minimal.

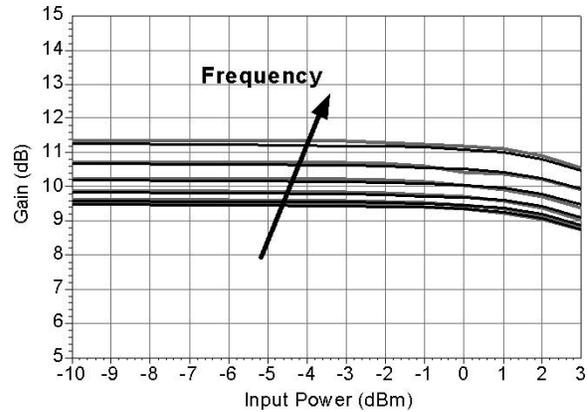


Fig. 6. Single-tone power sweep comparing the gain compression of the NLTA behavioral model (gray line) and the transistor-level circuit model (black). The frequency range is 1–11 GHz.

This indicated that the network had approximated the function, but was not yet overtrained. The mean square errors in the function fitting of all the data were of the order  $10^{-4}$ .

The neural-network models were exported from MATLAB and implemented as an SDD in ADS: the SDD external port currents are defined by the neural-network function expressions. The embedding variables for the neural network are calculated by the SDD from the terminal voltages and currents.

The verification procedure in ADS included the following:

- 1) single-tone power sweep harmonic balance simulation, over 1–11-GHz frequency range and to at least  $P - 1$  dB; in addition to the power magnitude and phase responses, we observe the harmonic-distortion performance;
- 2) two-tone power-sweep simulation, over the same power and frequency range as 1), with a tone spacing of 100 MHz;
- 3) small-signal ( $S$ -parameter) frequency response;
- 4) transient simulation;
- 5) transient envelope simulation of adjacent channel leakage ratio (ACLR) using wide-band code-division multiple-access (WCDMA) input signal.

In addition to the above microwave performance of the behavioral model itself, we verified its performance of a cascade of amplifiers, thereby demonstrating the suitability of this approach for creating behavioral models for use in the simulation of large systems.

The power levels and frequencies used for validation were different from those used in the data/model generation. In addition, we investigated the limiting cases of linear or small-signal behavior using  $S$ -parameter simulation and dc behavior. Again, it is important to note that neither small-signal, nor dc data were used in the model generation procedure: only large-signal data were used.

In Fig. 6, the single-tone gain compression characteristic reproduced almost exactly by the behavioral model. The frequency range is 1–11 GHz, which is the operating bandwidth of the amplifier. In Fig. 7, phase is also faithfully reproduced.

In Fig. 8 we show the response up to the seventh harmonic for a single-tone input at 3 dBm, which is the  $P - 1$  dB compression point. There are some deviations, but this quality

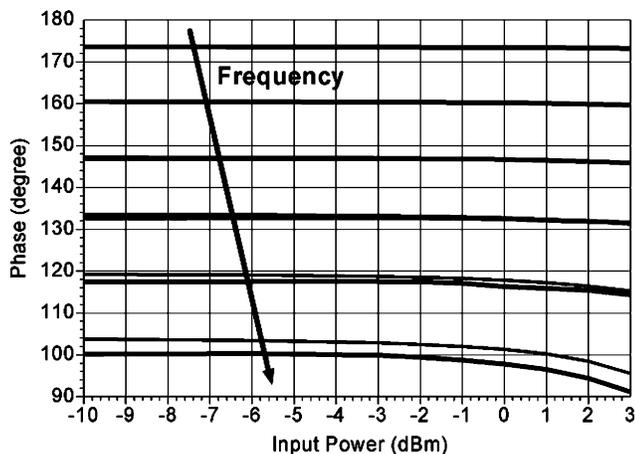


Fig. 7. Single-tone power sweep comparing the fundamental phase response of the NL TSA behavioral model (gray line) and the transistor-level circuit model (black). The frequency range is 1–11 GHz.

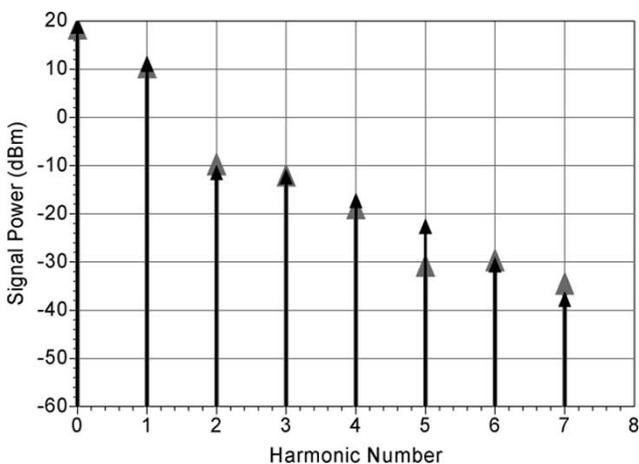


Fig. 8. Comparison of the dc and harmonic response in HB simulation for the NL TSA behavioral model (gray arrows) and the transistor-level circuit (black). This example is at  $P = 1$  dB compression and a fundamental frequency of 5 GHz in the middle of the amplifier passband.

of prediction is not obtained from simple models or heavily truncated—i.e., practical—Volterra-based models. The even harmonics are reproduced well here; this is not the case with simple “power out–power in” models, which *can only* predict odd-order harmonics. Second-order correctness is important, especially for “long-term” or slow memory effects, dc offsets, etc. Note also that the dc level is reproduced exactly by the behavioral model, even though no dc measurements were used in the model construction.

The model and circuit  $S$ -parameters are also in excellent agreement over the frequency range of 1–10 GHz, as shown in Fig. 9, indicating that the fully nonlinear model reduces to the correct linear behavior under small-signal conditions.

The two-tone performance of the behavioral model is also very accurate. This is shown in Fig. 10 for fundamental input signals of 2.0 and 2.1 GHz, at 0 dBm each tone, corresponding to approximately 1 dB of compression.

The time-domain output voltage waveforms for the two-tone input are shown in Fig. 11. The RF signal is modeled accurately, and the envelope signal at 100 MHz is also reproduced exactly.

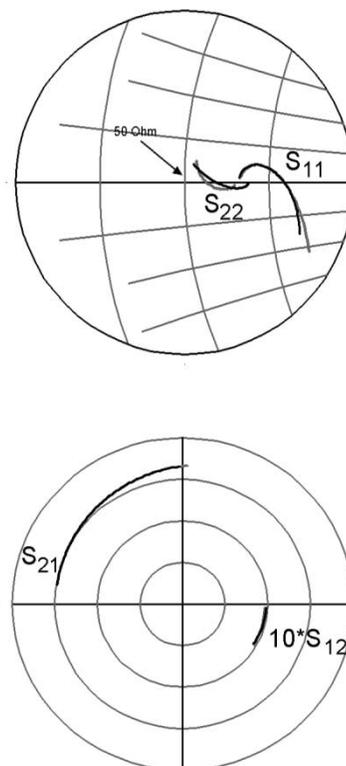


Fig. 9. Comparison of the  $S$ -parameters of the NL TSA behavioral model (gray lines) and the transistor-level circuit (black) over 1–10 GHz.

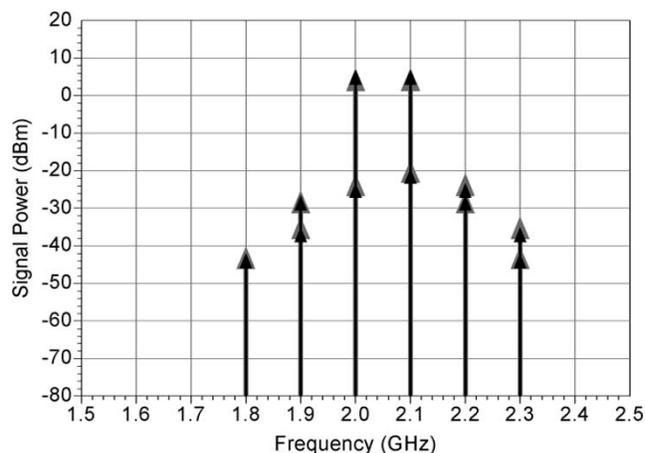


Fig. 10. Response to a two-tone excitation at 2.0 and 2.1 GHz, comparing the NL TSA behavioral model (gray arrows) and the transistor-level circuit (black). The input tones are each 0 dBm, which is approximately 1-dB compression.

Again, this is an excellent performance as no low-frequency IF signals were used in the creation of the behavioral model.

In the above verification, the behavioral model SDD and the full transistor-level circuit model were simulated in ADS. The simulation times for the circuit model and SDD model were found to be approximately comparable: the SDD model executed in approximately 25% less CPU time on a PC than the full circuit model. It is expected that a compiled model with a neural-network evaluation function would be significantly faster than the SDD implementation. Further, a more compact set of embedding variables could be chosen, and more compact neural-network structures can be achieved by using more sophisticated

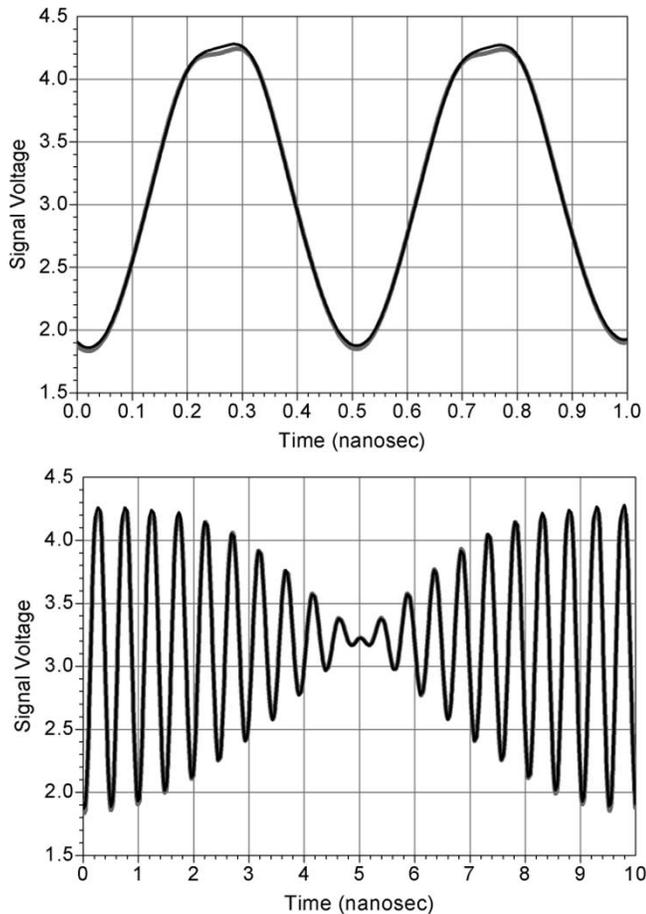


Fig. 11. Time-domain representation of the two-tone response shown in Fig. 10 obtained by FFT. The NL TSA behavioral model displays good fidelity with the transistor-level circuit response even at IF of 100 MHz.

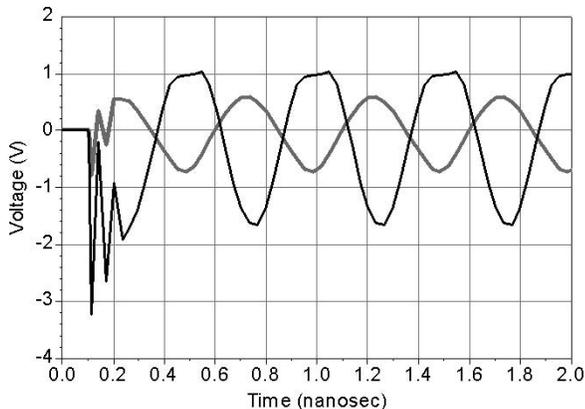


Fig. 12. Transient response of the NL TSA behavioral model, the input is shown dotted (gray line) and the output is shown black. The model shows rapid convergence even at a signal level of +5 dBm, well into saturation for this amplifier and much higher than the level of the training signals.

training and pruning algorithms. Both strategies will also improve simulation speed.

The NL TSA behavioral model operates successfully in Transient Simulation (Fig. 12) as well as harmonic balance. The input signal level here is very high—above the  $P - 1$  dB test

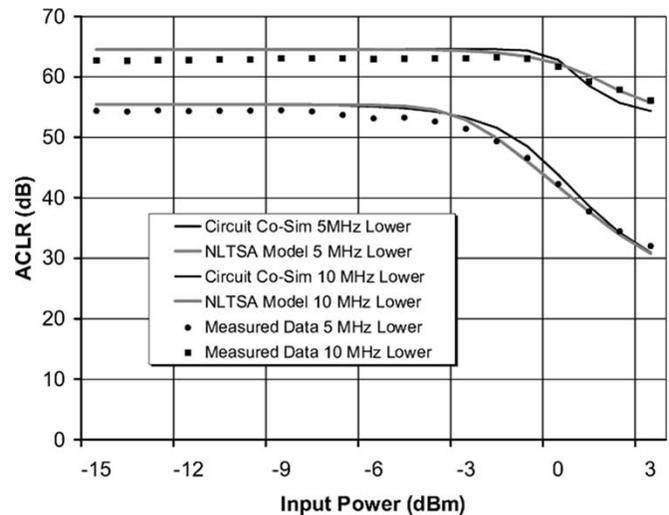


Fig. 13. Comparison of the NL TSA behavioral model and the transistor-level circuit in transient envelope simulation excited by a WCDMA-modulated signal at 2 GHz for ACLR prediction. The measured data are also included. The models show good agreement with each other and the measurements.

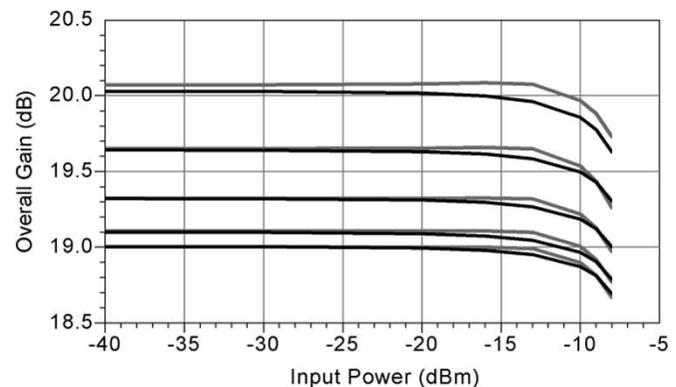


Fig. 14. Comparison of the gain compression characteristic for a cascade of two circuits: the cascade of NL TSA behavioral models (gray line) and the transistor-level circuits (black). Excellent agreement is obtained, noting the expanded ordinate scale.

level. The model is predicting hard limiting behavior correctly. This is often difficult for dynamical models to predict well, especially Volterra models.

Fig. 13 shows a comparison of the behavioral model and the transistor level circuit in a transient envelope simulation in ADS Ptolemy. The excitation is a WCDMA modulated signal at 2 GHz. Similar results have been obtained for error vector magnitude (EVM) using wireless local area network (WLAN) standard signals. This aspect of the model performance is particularly good since no digital modulation signals were used in the model generation process, only sinusoidal signals were used. In this case, the SDD model executed significantly faster than the circuit model: 294 CPU s compared with 1532 CPU s for a single power point.

While the accurate agreement between the behavioral model results and those from the transistor-level circuit are an essential first step in validating the behavioral model, the usefulness of this model is pertinent in the simulation of a module or system

containing several components. To demonstrate that the behavioral model of the amplifier can be used in a system-level simulation, we place two models in cascade, and compare the results with two transistor-level circuit models in cascade. The simulation results for the gain-compression characteristic are shown in Fig. 14. Excellent agreement between the behavioral model and transistor-level circuit is observed. Similar results are also found for the harmonic performance, IM behavior, etc. of the cascade.

## VII. CONCLUSIONS

We have presented a new, general, and systematic time-domain methodology for generating nonlinear behavioral models based on well-established techniques from nonlinear dynamics, system identification, and computational geometry. The modeling technique we have described is general, systematic, and scalable. The order of the model is contained in the embedding dimension and the ANN structure, and is vastly smaller than the number of internal degrees of freedom of the DUT. This model-order reduction is achieved through a systematic procedure, and does not require detailed knowledge of the circuit of the DUT, or skilled analysis of the circuit equations necessary for traditional model-order reduction methods.

A prototype Behavioral Modeling Toolbox has been developed in MATLAB, which reads measured or simulated time-domain data and generates a model file that can be imported into the Agilent-EEsof ADS nonlinear microwave circuit simulator. With this toolbox, we have generated a behavioral model from simulated data using a transistor-level circuit model of a broad-band microwave IC amplifier. The behavioral model faithfully reproduces the circuit model electrical behavior in a wide range of validation exercises including single- and two-tone power-frequency sweeps over the operating space of the amplifier, dc conditions, and  $S$ -parameter simulation. The cascading of two microwave amplifiers is also modeled accurately, indicating that these behavioral models can be used in system-level simulations of modules containing several amplifiers.

## ACKNOWLEDGMENT

The authors thank G. Jue, Agilent Technologies Inc., Liberty Lake, WA, for performing the transient envelope simulations and ACLR measurement shown in Fig. 13.

## REFERENCES

- [1] K. S. Kundert and A. Sangiovanni-Vincentelli, "Simulation of nonlinear circuits in the frequency domain," *IEEE Trans. Computer-Aided Design*, vol. CAD-5, pp. 521–535, Oct. 1986.
- [2] D. Sharrit, "A new method of analysis of communication systems," presented at the IEEE MTT-S Int. Microwave Symp. Workshop, 1996.
- [3] E. Ngoya and R. Larcheveque, "Envelope transient analysis: A new method for the transient and steady state analysis of microwave communication circuits and systems," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 1996, pp. 1365–1368.
- [4] J. Phillips, "A statistical perspective on nonlinear model reduction," in *Behavioral Modeling and Simulation Workshop*, San Jose, CA, Oct. 7–8, 2003, pp. 41–46.
- [5] D. M. Walker, R. Brown, and N. B. Tuffillaro, "Constructing transportable behavioral models for nonlinear electronic devices," *Phys. Lett. A*, vol. 255, pp. 236–242, 1999.
- [6] D. E. Root, J. Wood, and N. B. Tuffillaro, "New techniques for nonlinear behavioral modeling of microwave/RF IC's from simulation and nonlinear microwave measurements," in *Proc. Design Automation Conf.*, 2003, pp. 85–90.
- [7] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, MA: Cambridge Univ. Press, 1997.
- [8] M. Casdagli, "A dynamical systems approach to modeling input–output systems," in *Nonlinear Modeling and Forecasting, SFI Studies in the Science of Complexity*, M. Casdagli and S. Eubank, Eds. Reading, MA: Addison-Wesley, 1992, vol. XII.
- [9] J. Stark, "Delay embeddings and forced systems," *J. Nonlinear Sci.*, vol. 9, pp. 255–332, 1999.
- [10] M. Rewienski and J. White, "A trajectory piecewise-linear approach to model order reduction and fast simulation of nonlinear circuits and micro-machined devices," *IEEE Trans. Computer-Aided Design*, vol. 22, pp. 155–170, Feb. 2003.
- [11] D. M. Walker and N. B. Tuffillaro, "Phase space reconstruction using input–output time series data," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 60, pp. 4008–4013, 1999.
- [12] N. Gershenfeld, *The Nature of Mathematical Modeling*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [13] M. B. Kennel, R. Brown, and H. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A, Gen. Phys.*, vol. 45, pp. 3403–3411, 1992.
- [14] C. Rhodes and M. Morari, "False-nearest-neighbors algorithm and noise-corrupted time-series," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 55, pp. 6162–6170, 1997.
- [15] J. Xu, M. C. E. Yagoub, R. Ding, and Q.-J. Zhang, "Neural-based dynamic modeling of nonlinear microwave circuits," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 2002, pp. 1101–1104.
- [16] D. Schreurs, J. Wood, N. B. Tuffillaro, D. Usikov, L. Barford, and D. E. Root, "The construction and evaluation of behavioral models for microwave devices based on time-domain large-signal measurements," in *Int. Electron Devices Meeting Conf. Dig.*, 2000, pp. 819–822.
- [17] D. M. Walker, N. B. Tuffillaro, and P. Gross, "Radial basis model for feedback systems with fading memory," *IEEE Trans. Circuits Syst. I*, vol. 48, pp. 1147–1151, Sept. 2001.
- [18] D. Schreurs, J. Wood, N. B. Tuffillaro, L. Barford, and D. E. Root, "Construction of behavioral models for microwave devices from time-domain large signal measurements to speed up high-level design simulations," *J. RF Microwave Computer-Aided Eng.*, vol. 13, pp. 54–61, 2003.
- [19] J. Wood and D. E. Root, "The behavioral modeling of microwave/RF IC's using nonlinear time series analysis," in *IEEE MTT-S Int. Microwave Symp. Dig.*, 2003, pp. 791–794.
- [20] G. Cybenko, "Approximation by superposition of sigmoidal functions," *Math. Contr., Signals, Syst.*, vol. 2, pp. 303–314, 1989.
- [21] S. Haykin, *Neural Networks*. Upper Saddle River, NJ: Prentice-Hall, 1999.
- [22] L. Prechelt, "Automatic early-stopping using cross validation: Quantifying the criteria," *Neural Networks*, vol. 11, pp. 761–767, 1998.
- [23] F. Dan Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proc. Int. Joint Neural Networks Conf.*, 1997, pp. 1930–1935.
- [24] D. E. Root, J. Wood, A. Pekker, N. Tuffillaro, and D. Schreurs, "Systematic behavioral modeling of nonlinear microwave/RF circuits in the time domain using techniques from nonlinear dynamical systems," presented at the Behavioral Modeling and Simulation Workshop, Santa Rosa, CA, Oct. 6–8, 2002.
- [25] D. Barataud, A. Mallet, M. Campovecchio, J. M. Nebus, J. P. Vilotte, and J. Verspecht, "Measurements of time-domain voltage/current waveforms at RF and microwave frequencies for the characterization of nonlinear devices," in *Instrument and Measurement Technology Conf.*, vol. 2, 1998, pp. 1006–10.
- [26] K. Judd and A. Mees, "On selecting models for nonlinear time series," *Physica*, vol. D 82, pp. 426–444, 1995.
- [27] M. Small and C. K. Tse, "Minimum description length neural networks for time-series prediction," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 66, Paper 066 701, pp. 1–12, 2002.
- [28] HMMC-5200 DC–20 GHz HBT series-shunt amplifier datasheet. Agilent Technol., Palo Alto, CA. [Online]. Available: [http://www.agilent.com/home/moved/spg\\_moved.shtml](http://www.agilent.com/home/moved/spg_moved.shtml)



**John Wood** (M'87–SM'03) received the B.Sc. and Ph.D. degrees in electrical and electronic engineering from The University of Leeds, Leeds, U.K., in 1976 and 1980, respectively.

From 1983 to 1997, he was a member of the academic staff with the University of York, York, U.K., where he was responsible for teaching and research in solid-state electronics and microwave device and circuit technology. In 1997, he joined the Microwave Technology Center, Agilent Technologies Inc. (formerly the Hewlett-Packard Company), Santa Rosa,

CA. He is currently with the Computer-Aided Engineering, Modeling, and Advanced Characterization Group. He has authored or coauthored approximately 70 papers and articles. His recent research has included the investigation and development of analytic large-signal field-effect transistor (FET) models, and bias-dependent linear FET models for millimeter-wave applications, HBT modeling, and nonlinear behavioral modeling using LSNA measurements, and nonlinear system identification techniques.



**David E. Root** (M'89–SM'01–F'02) received the B.S. degrees in physics and mathematics and Ph.D. degree in theoretical physics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1986, respectively.

He is currently Principal Research Scientist with the Worldwide Process and Technology Centers, Agilent Technologies Inc. (formerly the Hewlett-Packard Company), Santa Rosa, CA. Since joining the Hewlett-Packard Company in 1985, he has held numerous scientific and management positions. He has authored or coauthored approximately 70 technical publications, primarily in the areas of nonlinear device, behavioral, and statistical modeling and simulation. He originated and co-developed the commercial HP measurement-based large-signal HP field-effect transistor (HPFET) MESFET/high electron-mobility transistor (HEMT), MOSFET, and diode models, model generators, and associated automated data-acquisition systems. He co-originated, co-developed, and managed the creation of the Agilent HBT nonlinear simulation model for III–V HBTs. His current responsibilities include nonlinear device modeling, behavioral modeling, and techniques for combining nonlinear measurements, modeling, and simulation for new technical capabilities and business opportunities for Agilent Technologies Inc.

Dr. Root is a member of the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) Committee on CAD (MTT-S-1) and the Technical Program Committee of the IEEE MTT-S International Microwave Symposium (IMS). He is a reviewer for the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES.



**Nicholas B. Tuffillaro** is currently with Agilent Technologies Inc. (formerly the Hewlett-Packard Company), Palo Alto, CA, where he develops new modeling and test methods for RF, microwave, and optical devices. Prior to joining Agilent Technologies Inc., he performed experimental research on spatial temporal chaos in surface waves and theoretical research in the topological characterization of low-dimensional chaos. He has also been a consultant for DOW Chemical, the Ford Motor Company, and the Hewlett-Packard Company, during which

time he was involved with issues of nonlinear system identification, modeling, and analysis.