Dynamic Deviation Reduction-Based Volterra Behavioral Modeling of RF Power Amplifiers

Anding Zhu, *Member, IEEE*, José C. Pedro, *Senior Member, IEEE*, and Thomas J. Brazil, *Fellow, IEEE*

***Abstract*—A new representation of the Volterra series is proposed, which is derived from a previously introduced *modified Volterra series*, but adapted to the discrete time domain and reformulated in a novel way. Based on this representation, an efficient model-pruning approach, called *dynamic deviation reduction*, is introduced to simplify the structure of Volterra-series-based RF power amplifier behavioral models aimed at significantly reducing the complexity of the model, but without incurring loss of model fidelity. Both static nonlinearities and different orders of dynamic behavior can be separately identified and the proposed representation retains the important property of linearity with respect to series coefficients. This model can, therefore, be easily extracted directly from the measured time domain of input and output samples of an amplifier by employing simple linear system identification algorithms. A systematic mathematical derivation is presented, together with validation of the proposed method using both computer simulation and experiment.**

***Index Terms*—Behavioral model, power amplifiers (PAs), Volterra series.**

I. Introduction

In a wideband wireless system, the distortion induced by a power amplifier (PA) can be considered to arise from different sources or can be assigned to different physical phenomena such as: 1) static (device) nonlinearities; 2) linear memory effects, arising from time delays, or phase shifts, in the matching networks and the device/circuit elements used; and 3) nonlinear memory effects, such as those caused by trapping effects, nonideal bias networks, temperature dependence on the input power, etc. To accurately model a PA, we have to take into account both nonlinearities and memory effects. Although many behavioral models for RF PAs have been developed in recent years [1], this kind of modeling technique is still far from being mature since accurately characterizing PAs becomes more and more difficult and challenging as wireless systems migrate to higher frequencies, higher speeds, and wider bandwidths.

Manuscript received April 6, 2006; revised June 28, 2006. This work was supported by the Science Foundation Ireland under the Principal Investigator Award . This work was supported in part by the Network of Excellence TARGET under the 6th Framework Program funded by the European Commission.

A. Zhu and T. J. Brazil are with the School of Electrical, Electronic and Mechanical Engineering, University College Dublin, Dublin 4, Ireland (e-mail: anding.zhu@ucd.ie; tom.brazil@ucd.ie).

J. C. Pedro is with the Institute of Telecommunications, Universidade de Aveiro, 3810-193 Aveiro, Portugal (e-mail: jcpedro@det.ua.pt).

Color versions of Figs. 6–8 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMTT.2006.883243

The Volterra series provides a general way to model a nonlinear system with memory, and it has been used by several researchers to describe the relationship between the input and output of an amplifier [1]–[7]. However, high computational complexity continues to make methods of this kind rather impractical in some real applications because the number of parameters to be estimated increases exponentially with the degree of nonlinearity and with the memory length of the system. To overcome the high complexity of a general Volterra series, a Volterra-like approach, called the *modified Volterra series*, was proposed in [8] by Filicori and Vannini to model microwave transistors, and then extended to model PAs by Mirri *et al.* [9], [10] and Ngoya *et al.* [11]. This modified series has the important property that it separates the purely static effects from the dynamic ones, which are intimately mixed in the classical series. However, this modified Volterra series loses the property of linearity in model parameters, which means that the output of the model is no longer linear with respect to the coefficients [9]. This leads to the consequence that models of this kind cannot be extracted in a direct and systematic way using established linear system estimation procedures such as the least squares (LS) techniques, as is usual in the classical case. In fact, although the static part and different order dynamics can be estimated separately, extracting higher order dynamics involves complicated measurement procedures [9]–[11].

In this paper, we first extend the modified Volterra series to the discrete time domain, and rewrite it in the classical format after dynamic-order truncation. We then propose a new format of representation for the Volterra model, in which the input elements are organized according to the order of dynamics involved in the model. This is similar to the modified Volterra series, but retains the property of linearity in the parameters of the model, as for the classical Volterra series.

Based on this new representation, an effective model-order reduction method is proposed, called *dynamic deviation reduction* [12], in which higher order dynamics are removed since the effects of nonlinear dynamics tend to fade with increasing order in many real PAs. Unlike the classical Volterra model, where the number of coefficients increases exponentially with the nonlinearity order and memory length, in the proposed reduced-order model, the number of coefficients increases almost linearly with the order of nonlinearity and memory length. Since the model complexity is significantly reduced after dynamic-order truncation, this Volterra model can be used to accurately characterize a PA with static strong nonlinearities and with long-term linear and low-order nonlinear memory effects. Furthermore, the proposed model takes advantage of the properties of the modified Volterra series so that the static nonlinearities and different order

dynamics can be separated after model extraction, which provides us with an effective way to derive efficient distortion compensation approaches for PA linearization. Finally, since this model is built in the discrete time domain, it can be directly embedded in system-level simulation tools and implemented in digital circuits.

This paper is organized as follows. In Section II, after introducing the modified Volterra series, we present the new representation of the Volterra series. Based on this new representation, a dynamic model-order reduction is introduced in Section III. Model extraction procedures and experimental verification are given in Sections IV and V, respectively, with a conclusion in Section VI.

II. Volterra series

A Volterra series is a combination of linear convolution and a nonlinear power series so that it can be used to describe the input/output relationship of a general nonlinear, causal, and time-invariant system with fading memory. In the discrete time domain, a Volterra series can be written as

$$y(n)=\sum_{p=1}^{P}\sum_{i_1=0}^{M}\cdots\sum_{i_p=0}^{M}h_p(i_1,\ldots,i_p)\prod_{j=1}^{p}x(n-i_j) \quad (1)$$

where $x(n)$ and $y(n)$ represents the input and output, respectively, and $h_p(i_1,\ldots,i_p)$ is called the pth-order *Volterra kernel*. In real applications, and assumed in (1), the Volterra series is normally truncated to finite nonlinear order P and finite memory length M [2].

Unlike neural networks or other nonlinear functions, the output of the Volterra model is linear with respect to its coefficients. Under the assumption of stationarity, if we solve for the coefficients with respect to a minimum mean or least square error criterion, we will have a single global minimum. Therefore, it is possible to extract the nonlinear Volterra model in a direct way by using linear system identification algorithms. The Volterra series has been successfully used to solve many problems in science and engineering [2]–[4]. However, since all nonlinearities and memory effects are treated in the same way, the number of coefficients to be estimated increases exponentially with the degree of nonlinearity and with the memory length of the system. It is very difficult to identify a practically convenient experimental procedure for the measurement of kernels of order greater than five so that the described classical Volterra-series formulation can only be practically used for the characterization of weakly nonlinear systems.

A. Modified Volterra series

To overcome the limitation of the classical Volterra series, a Volterra-like approach, called a *modified Volterra series* [8]–[10], or *dynamic Volterra series* [11], was proposed, in which the input/output relationship for a nonlinear system with memory is described as a memoryless nonlinear term plus a purely dynamic contribution. This was based on introducing the dynamic deviation function $e(n,i)$

$$e(n,i)=x(n-i)-x(n) \quad (2)$$

which represents the deviation of the delayed input signal $x(n-i)$ with respect to the current input $x(n)$. Substituting (2) in (1), we obtain

$$\begin{aligned}y(n)&=\sum_{p=1}^{P}\sum_{i_1=0}^{M}\cdots\sum_{i_p=0}^{M}h_p(i_1,\ldots,i_p)\prod_{j=1}^{p}[x(n)+e(n,i_j)]\\&=\sum_{p=1}^{P}x^p(n)\sum_{i_1=0}^{M}\cdots\sum_{i_p=0}^{M}h_p(i_1,\ldots,i_p)\\&\quad+\sum_{p=1}^{P}\sum_{r=1}^{p}x^{p-r}(n)\binom{p}{r}\sum_{i_1=0}^{M}\cdots\\&\quad\sum_{i_p=0}^{M}h_p(i_1,\ldots,i_p)\prod_{j=1}^{r}e(n,i_j). \quad (3)\end{aligned}$$

After regrouping the coefficients, it is immediately apparent that the output signal $y(n)$ in (1) can be expressed through the following dynamic-deviation-based Volterra-like series:

$$y(n)=y_s(n)+y_d(n) \quad (4)$$

where $y_s(n)$ and $y_d(n)$ represents the static and dynamic characteristics of the system, respectively.

$y_s(n)$ can be expressed as a power series of the current input signal $x(n)$

$$y_s(n)=\sum_{p=1}^{P}a_px^p(n) \quad (5)$$

where a_p is the coefficient of the polynomial function, while $y_d(n)$ is the purely dynamic part

$$\begin{aligned}y_d(n)&=\sum_{p=1}^{P}\sum_{r=1}^{p}x^{p-r}(n)\\&\quad\cdot\sum_{i_1=0}^{M}\cdots\sum_{i_r=0}^{M}w_{p,r}(i_1,\ldots,i_r)\prod_{j=1}^{r}e(n,i_j) \quad (6)\end{aligned}$$

where $w_{p,r}(\cdot)$[1] represents the rth-order dynamic kernel of the pth-order nonlinearity. The relationship between the coefficients of the classical Volterra formulation $h_p(i_1,\ldots,i_p)$ and those of the dynamic series a_p and $w_{p,r}(\cdot)$ is presented in the Appendix.

In this model, the static nonlinearities and the dynamic part are separated. Furthermore, controlling the value of r, i.e., the order of the dynamics, allows us to truncate the model to a simpler version. For instance, in [8]–[11], (6) was truncated to first order. In that case, the modified Volterra model would have the form

$$y(n)=\sum_{p=1}^{P}a_px^p(n)+\sum_{p=1}^{P}x^{p-1}(n)\sum_{i=0}^{M}w_{p,1}(i)e(n,i) \quad (7)$$

in which only the static and first-order dynamic behavior are retained.

[1]Note that here we use $w_{p,r}(\cdot)$ instead of $g_p[x(n),\ldots]$, which was used in [10], to represent the dynamic coefficients. $g_p[x(n),\ldots]$ is the combination of $w_{p,r}(\cdot)$ and $x^r(n)$, which depends nonlinearly on $x(n)$, while $w_{p,r}(\ldots)$ is independent from $x(n)$.

However, the static part and the different order dynamics have to be extracted separately in this model, which involves very complicated measurement procedures [10], [11], especially when higher order dynamics are included.

B. New Representation of Volterra Series

In order to take advantage of the modified Volterra series, but also keep the model extraction as simple as possible, we derive a new representation of the Volterra series here.

Let us start from the first-order truncated modified Volterra series in (7). At first sight, this modified Volterra model seems to be fundamentally different from the classical Volterra series. However, if we re-substitute (2) in (7), we obtain

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} a_p x^p(n) \\ &\quad + \sum_{p=1}^{P} x^{p-1}(n) \sum_{i=0}^{M} w_{p,1}(i)\left[x(n-i)-x(n)\right] \\ &= \sum_{p=1}^{P} a_p x^p(n) \\ &\quad + \sum_{p=1}^{P}\sum_{i=0}^{M} w_{p,1}(i)\left[x^{p-1}(n)x(n-i)-x^p(n)\right]. \end{aligned} \tag{8}$$

After some rearrangements, it can be shown that

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} h_{p,0}(0,\ldots,0)x^p(n) \\ &\quad + \sum_{p=1}^{P} x^{p-1}(n) \sum_{i=1}^{M} h_{p,1}(0,\ldots,0,i)x(n-i). \end{aligned} \tag{9}$$

Now, we truncate (6) to second order, and the model becomes

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} a_p x^p(n) + \sum_{p=1}^{P} x^{p-1}(n) \sum_{i=0}^{M} w_{p,1}(i)e(n,i) \\ &\quad + \sum_{p=1}^{P} x^{p-2}(n) \sum_{i_1=0}^{M}\sum_{i_2=0}^{M} w_{p,2}(i_1,i_2)e(n,i_1)e(n,i_2). \end{aligned} \tag{10}$$

In the same way, re-substituting (2) in (10), we obtain

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} a_p x^p(n) \\ &\quad + \sum_{p=1}^{P} x^{p-1}(n) \sum_{i=0}^{M} w_{p,1}(i)\left[x(n-i)-x(n)\right] \\ &\quad + \sum_{p=1}^{P} x^{p-2}(n) \sum_{i_1=0}^{M}\sum_{i_2=0}^{M} w_{p,2}(i_1,i_2) \\ &\quad \times \left[x(n-i_1)-x(n)\right]\left[x(n-i_2)-x(n)\right] \\ &= \sum_{p=1}^{P} a_p x^p(n) \\ &\quad + \sum_{p=1}^{P}\sum_{i=0}^{M} w_{p,1}(i)\left[x^{p-1}(n)x(n-i)-x^p(n)\right] \\ &\quad + \sum_{p=1}^{P}\sum_{i_1=0}^{M}\sum_{i_2=0}^{M} w_{p,2}(i_1,i_2) \\ &\quad \times \left[x^{p-2}(n)x(n-i_1)x(n-i_2)-x^{p-1}(n)x(n-i_1)\right. \\ &\quad \left. -x^{p-1}(n)x(n-i_2)+x^p(n)\right]. \end{aligned} \tag{11}$$

Regrouping the coefficients a_p, $w_{p,1}(\cdot)$, and $w_{p,2}(\cdot)$, we can write (11) as

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} h_{p,0}(0,\ldots,0)x^p(n) \\ &\quad + \sum_{p=1}^{P}\left[x^{p-1}(n)\sum_{i=1}^{M} h_{p,1}(0,\ldots,0,i)x(n-i)\right] \\ &\quad + \sum_{p=2}^{P}\left[x^{p-2}(n)\sum_{i_1=1}^{M}\sum_{i_2=i_1}^{M}\cdot h_{p,2}(0,\ldots,0,i_1,i_2)\right. \\ &\quad \left. \times x(n-i_1)x(n-i_2)\right]. \end{aligned} \tag{12}$$

Following the same procedures of (8)–(12), we can rewrite the classical Volterra series in (1) as

$$\begin{aligned} &y(n) \\ &= \sum_{p=1}^{P} h_{p,0}(0,\ldots,0)x^p(n) \\ &\quad + \sum_{p=1}^{P}\left\{\sum_{r=1}^{p}\left[x^{p-r}(n)\sum_{i_1=1}^{M}\cdots\sum_{i_r=i_{r-1}}^{M}\right.\right. \\ &\quad \left.\left. \cdot h_{p,r}(0,\ldots,0,i_1,\ldots,i_r)\prod_{j=1}^{r}x(n-i_j)\right]\right\} \end{aligned} \tag{13}$$

which leads to a new representation of Volterra series, where $h_{p,r}(0,\ldots,0,i_1,\ldots,i_r)$ represents the pth-order Volterra kernel where the first $p-r$ indices are "0," corresponding to the input item $x^{p-r}(n)x(n-i_1),\ldots,x(n-i_r)$. Compared to (1), we can see that the coefficients $h_{p,r}(\cdot,\ldots,\cdot)$ in (13) are the same as in the classical expression, but the sequence of the input product items in the input vector has been changed. In this new representation, r represents the possible number of product terms of the delayed inputs in the input items. For instance, $r=1$ means only one delayed input is included in the product, i.e., $x^{p-1}(n)x(n-i)$. On the other hand, compared to the dynamic-deviation based series in (6), we see that r can also be interpreted as the order of the dynamics involved in the model since it decides how many deviations, i.e., $e(n,i_j)$, will be included in the input vector.

To make this clear, we now further demonstrate the properties of the new Volterra representation by using kernels indices with their corresponding input items, as shown in Table I, where a "0" corresponds to the current instantaneous input $x(n)$ and a non-"0", i.e., an "i," corresponds to the delayed input $x(n-i)$.

TABLE I
INDICES OF COEFFICIENTS

	r=0 (instantaneous)	r=1 (1st-order dynamics)	r=2 (2nd-order dynamics)	r=3 (3rd-order dynamics)	…
$h_{1,r}(\cdot)$	0	1, 2, 3, …, M			
$h_{2,r}(\cdot,\cdot)$	00	01, 02, 03,…, 0M	11, 12, …, 1M, 22, 23, …, MM		
$h_{3,r}(\cdot,\cdot,\cdot)$	000	001, 002, 003,…, 00M	011, 012, 013, …, 01M, 022, 023, …,033, 034, …, 0MM	111, 112, 113, …, 11M, 122, 123, …, 133, 134, …, MMM	…
…	…	…	…	…	…

In Table I, each row includes kernels from the same order of nonlinearity, while the columns are divided by the order of the dynamics involved in the model. We can easily see that the value of r directly indicates how many non-"0" indices are in the kernel and, thus, find out how many delayed inputs are involved in the input products. For instance, $r = 2$ means that all its input products have two components from delayed inputs, e.g., "012" corresponds to $x(n)x(n-1)x(n-2)$. In the classical Volterra series of (1), the coefficients are gathered by rows in Table I, which is based on the orders of nonlinearity, while in the new representation of (13), the coefficients are organized by columns, i.e., the orders of the dynamics involved, where $r = 0$ corresponds to the first term in (13), then the second column, and so on. This makes the new Volterra model have the same advantages as the modified Volterra series. However, at the same time, the property of linearity in the coefficients is also kept from the classical Volterra series.

Thus, this new representation makes possible an effective model-order reduction method and a systematic distortion evaluation approach, while keeping the conventional linear procedures for the model extraction, as will be discussed in Sections III and IV.

III. DYNAMIC DEVIATION REDUCTION

In most real PAs, the distortions mainly arise from the static nonlinearities, and the effects of nonlinear dynamics in the PA fade with increasing order. This means that the static nonlinearities and low-order dynamics are the dominant sources of the distortions induced by the PA. Therefore, it is reasonable to remove higher order dynamics in the model to reduce the model complexity.

In [8]–[11], the modified Volterra series was truncated to the first order, in which only first-order dynamics are accounted for in the model. However, as discussed in [13], while this first-order truncation permits accurate modeling of highly nonlinear systems, its effectiveness tends to be limited to those systems where the nonlinear dynamics are sufficiently small so that they can be omitted. Unfortunately, it is found in practice that many solid-state amplifiers exhibit nonnegligible nonlinear dynamics, especially due to thermal and bias circuit modulation effects, which implies that nonlinear memory effects become apparent. In that situation, a first-order truncation is insufficient. Higher order dynamics must also be accounted for, and more terms need to be added to improve the accuracy of the model. However, we cannot simply add higher order terms to the dynamic model because increasing the order of dynamics in (6) results in a rapid growth of model extraction effort, requiring complicated multidimensional measurements.

Fortunately, the new Volterra representation in (13) provides us with a very flexible way to prune the Volterra model efficiently, keeping, at the same time, the model extraction complexity to an acceptable level, as is explained below.

Since in (13) r represents the order of the dynamics of the input products, we can easily control the order of dynamic behavior by limiting the value of r, i.e., setting $1 \leq r \leq R$, where R is a small number, thus pruning the model, as in the modified Volterra series. For instance, it is easy to see that the second-order truncated modified Volterra model in (10) is equivalent to the truncation of the new Volterra model in (13) by limiting $1 \leq r \leq 2$, which leads to (12).

The key difference between (10) and (12) is that they require different model extraction procedures. Extracting the model in (10) requires difficult experimental measurements, as described in [10] and [11], while the coefficients in (12) can be extracted in a direct way since the output of this model is still linear with respect to all coefficients. This can be done by using LS algorithms to estimate the Volterra kernels from measured arbitrary input and output data of the PA in the discrete time domain, as presented in Section IV.

The decision on how to select the truncation order depends on the practical characteristics of PAs and the model fidelity required. We refer to this model reduction approach as "dynamic deviation reduction" since r represents the order of the dynamic deviation in the model. Note that this dynamic-order truncation does not affect the nonlinearity or memory truncation in the same way as in the classical series. In other words, it only removes higher order dynamics, preserving the static nonlinearities and low-order dynamics. For example, in (12), we removed the higher order dynamics, whose order is over two—i.e., all kernels in column 3 and beyond in Table I were omitted—but we are still able to characterize high-order static nonlinearities by increasing P, or to model longer-term linear and first/second-order nonlinear dynamics by increasing M. In conclusion, we have three truncation parameters, i.e., P, M, and r, to choose from in this dynamic model-order reduction, which renders the application of the Volterra model much more flexible.

In the classical Volterra series, we only truncate the model by limiting the order of nonlinearity P and memory length M so

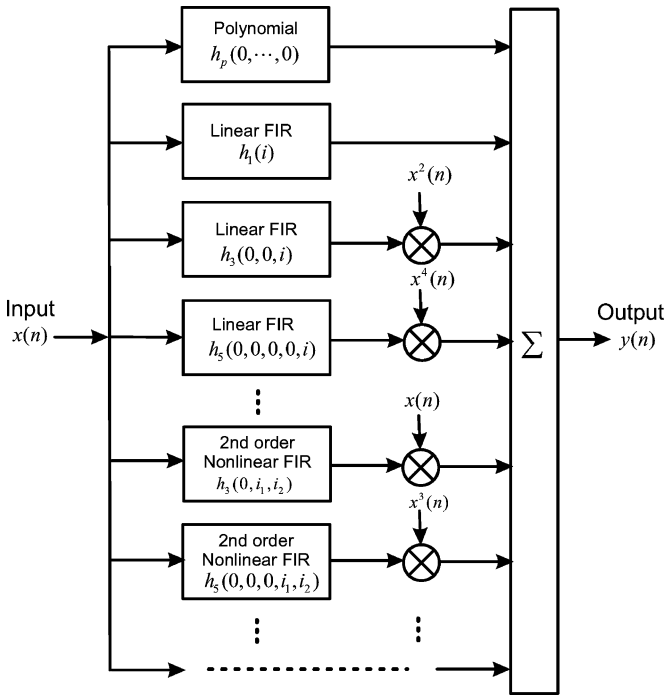

Fig. 1. Truncated Volterra model structure.

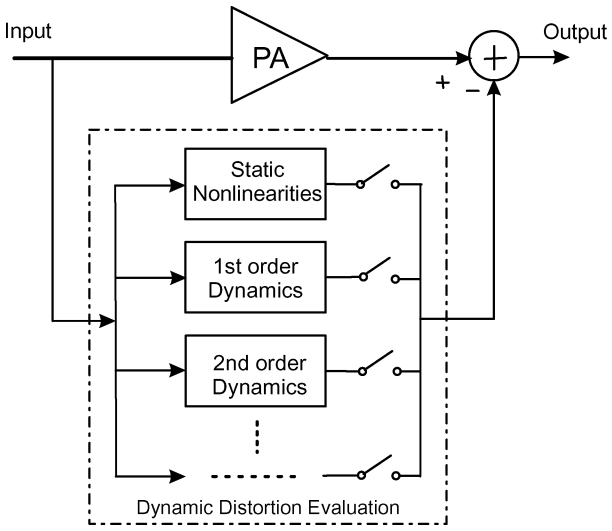

Fig. 2. Functional block diagram for dynamic distortion evaluation.

that the number of the pth-order coefficients is M^P.[2] However, after dynamic deviation reduction, the number of coefficients will only be $1+M+M^2+\cdots+M^r$.[3] When r has a small value, the total number of kernels can be kept reasonably small even with large values of P and M. This significantly simplifies the model structure and reduces the model extraction effort. However, model fidelity can still be acceptable since higher order dynamics do not significantly impact on the output of the PA.

Model implementation also becomes much simpler since only a limited number of multiplier products are needed, as shown in Fig. 1. There, a polynomial series is used to construct the instantaneous transfer function, while transversal finite impulse response (FIR) filters are used to implement the time shifts and convolutions [1], [3]. Note that, in this figure, the even-order nonlinear terms were omitted since only odd-order nonlinearities affect the first-zone output, i.e., the one where the information is transmitted. Also, only real RF signals were considered. For handling carrier-modulated signals, a low-pass equivalent Volterra model was developed in [12], where complex envelope signals were assumed.

Furthermore, as discussed in Section II, the new representation of the Volterra series in (13) is equivalent to the modified Volterra series of (4). This means that the coefficients of the modified Volterra series can be directly calculated from the coefficients extracted for (13) or vice versa. Therefore, when the Volterra kernels in (13) are extracted, the modified series in (4) can be easily constructed. The purely static effects and different orders of dynamics can then be separated. As illustrated in Fig. 2, this provides us with an effective way to investigate the origins of different kinds of distortion and to evaluate their effects on the output of the PA.

As the functional block diagram shows, the nonlinear static and different dynamic characteristics are built into several sub-blocks according to their orders. Hence, by switching on branches containing these blocks, we can observe how the distortion changes in the output, and can thereby evaluate the effects induced by the static or dynamic nonlinearities. Of course, similar blocks could be used upstream of the PA to operate as pre-distorters so as to cancel the distortions induced by the PA, although that is not detailed in this study.

IV. Model Extraction

A. Excitation Signals

In this study, we use arbitrary signals as the excitation. To make sure the excitation source is sufficiently rich to excite all important properties of the system, most of the techniques used for the extraction of the Volterra model to date are formulated through a system approach, using the correlation properties of Gaussian white noise [2]–[4]. However, it is not convenient to use these approaches with contemporary commercial circuit simulators or experimental measurements. This is because most commercial simulators cannot handle a general Gaussian white noise source, and also practical PAs are not excited by Gaussian noise signals. Thus, here, we use a combination of several time-domain wideband code division multiple access (W-CDMA) user signals to extract the Volterra kernels. W-CDMA is a popular air interface technology for third-generation RF cellular communication systems, which supports wide RF bandwidths, typically from 5 to 20 MHz. It is based on the direct-sequence code division multiple access (DS-CDMA) technique, i.e., user information bits are spread over a wide bandwidth by multiplying the user data with quasi-random bits (called chips) derived from code division multiple access

[2]For formulation simplicity, this number includes all kernels. If kernel symmetry is considered and omitting even-order kernels that do not contribute to the PA's first zone output, the total number of coefficients can be further reduced, but it still increases exponentially with P.

[3]This number can also be decreased if kernel symmetry is again considered and even orders are omitted.

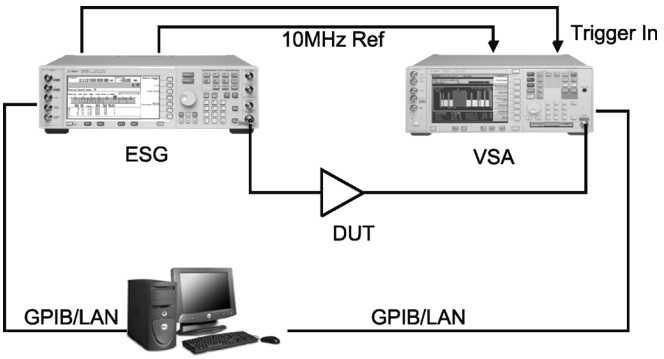

Fig. 3. Experimental test bench.

(CDMA) uncorrelated spreading codes. It is well known that a general model of a DS-CDMA system with spread-spectrum (SS) signals is described as [15]

$$s(t) = \sum_{i=1}^{n} m_i(t) c_i(t) \cos\left[2\pi f_0 t + \theta_i(t)\right] \tag{14}$$

where $m_i(t) = i\text{th}$ baseband quadrature or binary phase-shift keying modulated signal and $c_i(t) = i\text{th}$ pseudonoise binary code with a bandwidth of B. f_0 is the carrier frequency and $\theta_i(t)$ is the phase of the carrier associated with the ith SS signal. According to the law of large numbers and the central limit theorem in statistics, no matter what the distribution of each SS signal is, as n becomes large, the CDMA signal $s(t)$ will tend towards a (band-limited) zero-mean Gaussian stochastic process [15]. Therefore, a composite baseband CDMA or W-CDMA excitation signal can be utilized as an equivalent band-limited white Gaussian process to estimate the Volterra transfer function of a PA.

Furthermore, a W-CDMA signal has much higher peak-to-average power ratio and much wider modulation frequency components than the sometimes used two-tone signal, which means that it can drive the PA through a wider nonlinearity and dynamics region. W-CDMA signal sources are currently available in most commercial computer-aided design (CAD) software, e.g., Agilent ADS, MATLAB/Simulink, etc. It is also a built-in feature of most of the latest signal generators.

B. Extraction Methodology

As mentioned earlier, the output of the new Volterra model is linear with respect to its coefficients. It is, therefore, possible to extract the nonlinear Volterra model in a direct way by using arbitrary sampled input and output signals.

Recently, a time-domain stimulus-response measurement solution has been proposed by Agilent Technologies [16] (shown in Fig. 3), which uses arbitrary waveforms, e.g., complex W-CDMA envelopes, as the input excitation. Similar configurations are also used by many other companies and researchers. In this test system, the modulated data files are first created at baseband, downloaded to the arbitrary waveform generator, as complex in-phase (I) and quadrature (Q) signals, which are then fed to an IQ modulator present in the electronic signal generator (ESG). The signal generator produces the test signal to the PA, i.e., our device-under-test (DUT). The output of the DUT is then down-converted and sampled by the vector signal analyzer (VSA). The sampled input and output data are captured and finally used to extract behavioral models for the PA.

From the point-of-view of system identification, we can consider that the coefficients appearing in (13) are a generalization of the impulse response coefficients $h_{p,r}(\ldots)$ defining a linear model. Consequently, one possible approach to the problem of the Volterra model parameter estimation is to treat it as a large, but standard, linear regression problem. In particular, we could form a single large parameter vector $\boldsymbol{\theta}$ containing all of the unknown coefficients $h_{p,r}(\ldots)$ and define the matrix $\mathbf{X}$ including all of the product terms $x(n-i_1), \ldots, x(n-i_M)$ appearing in the model for $n = M+1, \ldots, N$, where N is the total length of the available data record. If we assume the presence of an unmodeled error $\mathbf{e} = [e(M+1), \ldots, e(N)]^T$, the Volterra model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e} \tag{15}$$

where $\mathbf{y} = [y(M+1), \ldots, y(N)]^T$. A popular solution to this problem is the LS method, in which $\boldsymbol{\theta}$ is estimated as the value $\hat{\boldsymbol{\theta}}$ that minimizes the model error criterion

$$J(\boldsymbol{\theta}) = \sum_{n=M+1}^{N} e^2(n) = \mathbf{e}^T\mathbf{e} \tag{16}$$

where $(\cdot)^T$ represents the transpose.[4] A standard result states that such an estimate can be given by

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{17}$$

This result has the advantage of notational simplicity and general applicability. Obviously, other linear adaptive techniques, such as the recursive least squares (RLS) and the least mean squares (LMS) algorithms, could also be here employed to estimate the model parameters.

C. Model Fidelity Evaluation

The PA behavioral model presented in this study operates on baseband time-domain waveforms. To directly assess the predictive accuracy of the model, a very useful time-domain waveform metric, termed the normalized mean square error (NMSE) [17], can be employed. This verification metric is the total power of the error vector between the measured and modeled waveforms, normalized to the measured signal power, given explicitly by

$$\begin{aligned}\text{NMSE} &:= 10\log \\ &\times \left\{ \frac{\sum_{k=1}^{N}\left[\left(y_{I,k}^{\text{meas}} - y_{I,k}^{\text{mod}}\right)^2 + \left(y_{Q,k}^{\text{meas}} - y_{Q,k}^{\text{mod}}\right)^2\right]}{\sum_{k=1}^{N}\left[\left(y_{I,k}^{\text{meas}}\right)^2 + \left(y_{Q,k}^{\text{meas}}\right)^2\right]} \right\}\end{aligned} \tag{18}$$

[4]For the low-pass equivalent model, this transpose becomes the Hermitian transpose.

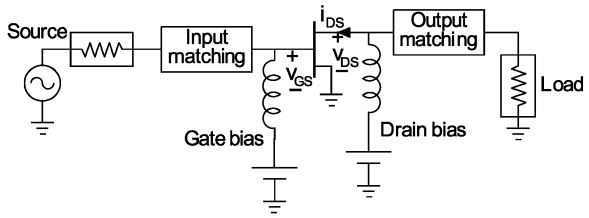

Fig. 4. Simplified schematic diagram of the simulated PA.

where the measured and modeled in-phase y_I and quadrature y_Q waveforms have N sample points. It is assumed that the "true" waveform is much closer to the measured waveform than the modeled waveform. Thus, the NMSE is indeed a metric of model fidelity.

However, many other system-level performance evaluation figures, e.g., adjacent channel power ratio (ACPR), error vector magnitude (EVM) and bit error ratio (BER), etc., could also be employed to evaluate the model fidelity.

V. Model Validation

Here, we verify the new behavioral model through both computer simulations and experimental tests.

The first example is intended to test the model ability in capturing the various nonlinear and dynamic effects of microwave PA circuits. Thus, we used a PA equivalent circuit in a standard microwave simulation software package to have easy control on the nonlinearity and memory effects. This also allowed us to eliminate noises and measurement errors in computer simulation, putting in evidence the actual model deficiencies. The disadvantage associated to such a test is that the validity of the behavioral model becomes obviously conditioned by the accuracy of the equivalent-circuit model used.

To make this modeling technique closer to the "real" world, we then also tested a commercial heterojunction bipolar transistor (HBT) PA in our laboratory. By using the Agilent connected-solution test bench shown in Fig. 3 [16], we captured the complex envelope data from the measured input and output of the PA, and then used them to extract and validate the behavioral model proposed.

Since only the envelopes carry useful information in these systems, all behavioral models herein extracted belong to the low-pass equivalent format [12].

A. Computer Simulations

In this test, we designed an equivalent-circuit PA model and simulated it with the Agilent ADS microwave simulation software package. This is a GaAs MESFET class-A PA operating at 2 GHz under a bias of 88% of $Idss$ with distributed matching networks. The block diagram is shown in Fig. 4. We used a W-CDMA signal as the excitation, and captured the simulated input and output data of the PA. These data were then used for model extraction and model validation.

To show the memory effects presented by the PA, we first simulated the circuit under a simple quadrature amplitude modulation (QAM) signal for various information bandwidths and for two different bias impedances. The resulting dynamic AM/AM plots are shown in Fig. 5 (the AM/PM plots have similar aspects, which are not shown here).

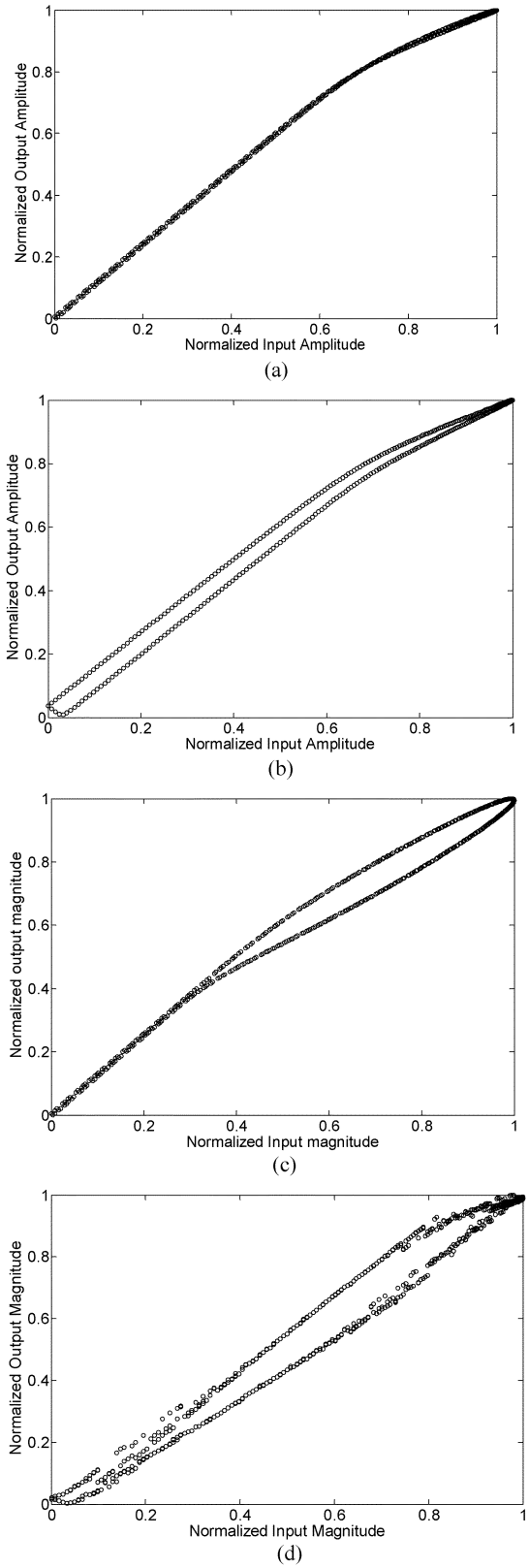

Fig. 5. Sample AM/AM diagrams indicate the memory effects presented in the PA. (a) 1-MHz envelope with ideal bias networks. (b) 10-MHz envelope with ideal bias networks. (c) 1-MHz envelope with nonideal bias networks. (d) 10-MHz envelope with nonideal bias networks.

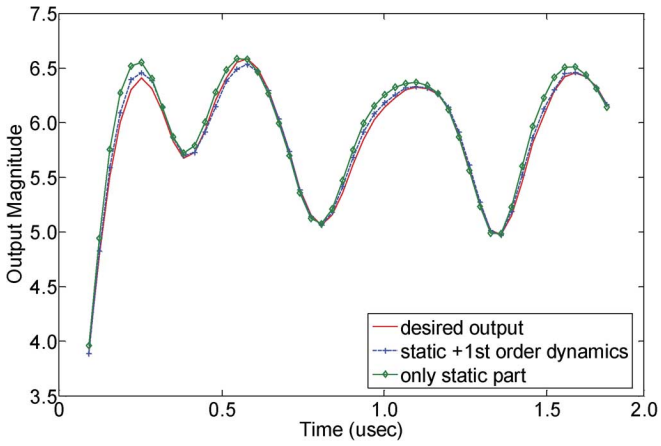

Fig. 6. Sample time-domain waveforms of predicted and simulated behavioral for the case (b) PA.

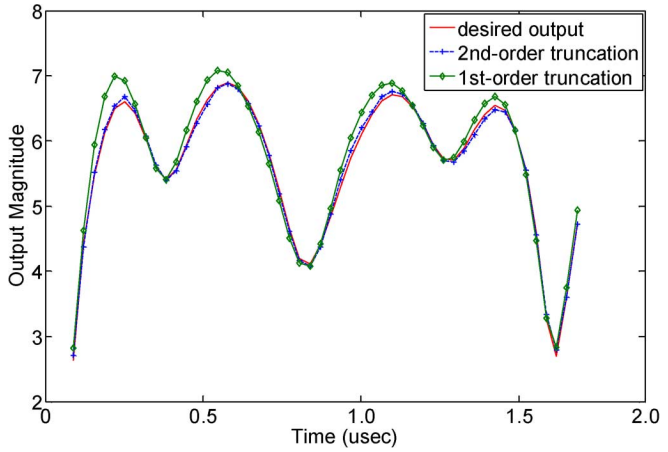

Fig. 7. Sample time-domain waveforms of predicted and simulated behavioral for the case (d) PA.

As shown in case (a), the PA is almost static when operated with a narrow band of 1 MHz and with ideal bias networks. Thus, in this case, a memoryless representation, such as the AM/AM and AM/PM model, could be used. However, when the bandwidth increases from 1 to 10 MHz, some memory effects become apparent (the AM/AM plot presents hysteresis), as shown in case (b). These memory effects are mainly due to the PA's matching networks, and they are mostly linear since they manifest themselves even in the PA small-signal region. Another widely known way to create memory effects for narrow bandwidths is to increase the reactance presented by the bias networks. That is shown in case (c), where now most of the effects are nonlinear, as can be seen from the fact that they only appear beyond the PA's onset of gain compression. Finally, in case (d) we have both linear and nonlinear memory effects (10-MHz bandwidth and for the increased bias impedance), something that could be expected from a real PA subject to wide bandwidth signals.

Since we want to evaluate the model's ability in treating various orders of system dynamics, we extracted a model for the PA as shown in case (b)—only first-order dynamics—and for the PA as shown in case (d)—first-order and higher order dynamics. In case (b), considering that the memory effects mainly emerged from linear and low-order nonlinear dynamics, we truncated the dynamic model to the first order, i.e., $r = 1$ with nonlinearity order $P = 5$ and memory length $M = 3$. The model was extracted via the LS estimation process proposed in Section IV. The average NMSE was -41 dB, a relative error less than 0.01%, which indicates that the first-order model predicted the PA output waveform quite well in this situation. We also calculated the coefficients for the equivalent modified Volterra model, and then the static and dynamic parts were separated. A sample of time-domain waveforms is shown in Fig. 6, where we can see that the model that includes only the static part is less accurate than the one including the dynamics. In case (d), a first-order $(r = 1)$ truncated model performed quite poorly as the NMSE could not get below -36 dB. However, when we added the second-order dynamics into the model, i.e., set $1 \leq r \leq 2$, the NMSE was improved to -42 dB. Again, the time-domain waveforms are shown in Fig. 7. These results clearly show that different orders of dynamic truncation have to be employed under different conditions, depending on the system characteristics and the desired model fidelity.

TABLE II
MODEL PERFORMANCE IN THE TIME DOMAIN

Order of Deviation	1	2	3	4	5
NMSE (dB)	-33.1	-37.2	-38.8	-38.1	-37.8
Number of Coefficients	18	54	118	184	244

B. Experimental Tests

In order to validate the proposed behavioral modeling technique in a real system, a commercial HBT class-AB PA was tested. This PA was operated at 2.14 GHz and excited by downlink 3GPP W-CDMA signals of 3.84-Mc/s chip rate and peak-to-average power ratio equal to 8.2 dB@0.01% probability on complementary cumulative distribution function (CCDF). The test bench setup used the ADS–ESG–VSA connected solution shown in Fig. 3. Around 12 000 sampling data points, with a sampling rate of 15.36 MHz, were captured from the PA input and output envelope signals.

In this test, the nonlinearity of the model was truncated to order 5, and the memory length was set to 3, i.e., $P = 5$ and $M = 3$. For comparison, we also truncated the order of the dynamics, i.e., set the value of r, from 1 to 5, which means that we extracted five different models from the first-order dynamic truncation to the full model. To evaluate the model's fidelity in the time domain, the NMSEs for each partial model were calculated. These results are shown in Table II, where we can see that the performance of the first-order model is quite poor. When the second-order dynamics were added in, the accuracy of the model was significantly improved. Nevertheless, increasing further the order of the dynamics only achieved minor improvements. The performance of the full model, in the case of maximum $r = 5$, was even worse than the truncated ones. This is common in nonorthogonal Volterra models, and is due to the fact that, when too many coefficients are involved, more uncertainties are brought into the model extraction process.

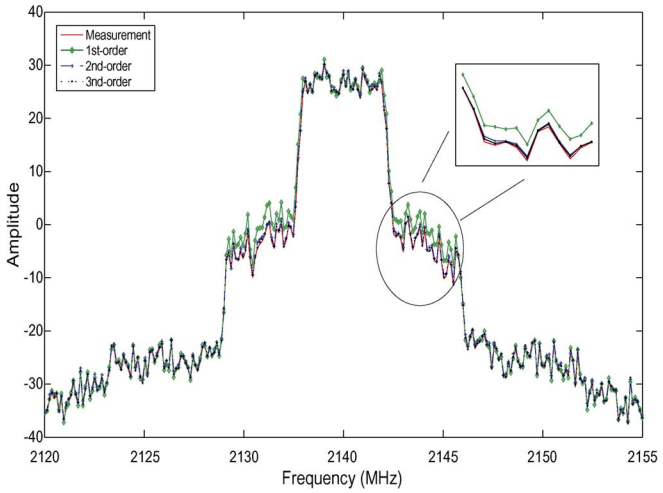

Fig. 8. Measured and modeled spectra of the PA first zone output.

The frequency-domain spectra of the outputs are shown in Fig. 8. From these results, we can see that the first-order dynamic model predicted the spectrum quite well in the in-band part, but some offsets appeared in the adjacent channels. The second-order model did better, but no obvious improvement was achieved after further increasing the order of the dynamics, which reflects what was already observed with the time-domain metric.

From these measurement results, we can conclude that static nonlinearities and low-order nonlinear dynamics do dominate nonlinear distortions caused by the tested PA. Therefore, it is indeed reasonable to remove higher order dynamics in the model, to reduce the model complexity, since their effects quickly fade with increasing order. Since the model structure becomes much simpler after model reduction, we gain room to increase the maximum order of nonlinearity P to cover higher order nonlinear effects, enabling in this way the application of the Volterra model to strongly nonlinear systems. In addition, we may increase the memory length M to characterize a wider range of long-term linear and low-order nonlinear memory effects if needed.

VI. Conclusion

A new format of Volterra series has been introduced in this paper, which consisted of regrouping the Volterra coefficients so that different dynamic orders can be controlled and separated, but keeping the easiness of the model extraction process. Based on this new representation, we proposed a "dynamic deviation reduction," which greatly simplified the model structure and, therefore, significantly reduced the complexity of Volterra-series-based behavioral models. Using this model reduction approach, we can effectively trade off between model simplicity and model fidelity, in a judicious manner, making the application of the Volterra model more flexible in practical applications. A model of this kind was shown to be easily extracted from time-domain measurements or simulations and simply implemented in system-level simulators.

Appendix

The relationship between the classical Volterra formulation and the dynamic Volterra series can be developed as follows.

In the static part, each coefficient a_p coincides with the multiple summation of the corresponding Volterra kernels of the same dimension

$$a_p = \sum_{i_1=0}^{M} \cdots \sum_{i_p=i_{p-1}}^{M} h_p(i_1, \ldots, i_p). \tag{A.1}$$

In the dynamic part, the coefficients of the first-order dynamics are

$$w_{1,1}(i) = h_1(i) \tag{A.2}$$

$$w_{2,1}(i) = h_2(0,i) + \underbrace{\sum^{M} h_2(i_1,i_2)}_{i_1=i \text{ or } i_2=i} \tag{A.3}$$

$$w_{3,1}(i) = h_3(0,0,i) + \underbrace{\sum^{M} h_3(0,i_1,i_2)}_{i_1=i \text{ or } i_2=i} + \underbrace{\sum^{M}\sum^{M} h_3(i_1,i_2,i_3)}_{i_1=i \text{ or } i_2=i \text{ or } i_3=i}$$

$$\vdots \tag{A.4}$$

Thus, we can see that $w_{p,1}(i)$ is the original pth-order Volterra kernel with the index $(0,\ldots,0,i)$ plus the sum of kernels of the same dimension with different indices, but in which one of them equals i.

The second-order coefficients are

$$w_{2,2}(i_1,i_2) = h_2(i_1,i_2) \tag{A.5}$$

$$w_{3,2}(i_1,i_2) = h_3(0,i_1,i_2) + \underbrace{\sum^{M} h_3(i_3,i_4,i_5)}_{if\ i_3,i_4,i_5 \in (i_1,i_2)} \tag{A.6}$$

$$w_{4,2}(i_1,i_2) = h_4(0,0,i_1,i_2) + \underbrace{\sum^{M} h_4(0,i_3,i_4,i_5)}_{if\ i_3,i_4,i_5 \in (i_1,i_2)} + \underbrace{\sum^{M}\sum^{M} h_4(i_3,i_4,i_5,i_6)}_{if\ i_3,i_4,i_5,i_6 \in (i_1,i_2)}$$

$$\vdots \tag{A.7}$$

The coefficients for higher order dynamics can be derived in the same way.

References

[1] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

[2] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, reprint ed. Melbourne, FL: Krieger, 1989.

[3] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*. New York: Wiley, 2000.

[4] V. Z. Marmarelis, *Nonlinear Dynamic Modeling of Physiological Systems*. New York: Wiley, 2004.

[5] A. Zhu, M. Wren, and T. J. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2003, pp. 787–790.

[6] A. Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned Volterra series," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, pp. 563–565, Dec. 2004.
[7] ——, "RF power amplifiers behavioral modeling using Volterra expansion with Laguerre functions," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2005, WE4D-1.
[8] F. Filicori and G. Vannini, "Mathematical approach to large-signal modeling of electron devices," *Electron. Lett.*, vol. 27, no. 4, pp. 357–359, 1991.
[9] D. Mirri *et al.*, "A modified Volterra series approach for nonlinear dynamic systems modeling," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 8, pp. 1118–1128, Aug. 2002.
[10] D. Mirri, F. Filicori, G. Iuculano, and G. Pasini, "A nonlinear dynamic model for performance analysis of large-signal amplifiers in communication systems," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 341–350, Apr. 2004.
[11] E. Ngoya *et al.*, "Accurate RF and microwave system level modeling of wideband nonlinear circuits," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2000, vol. 1, pp. 79–82.
[12] A. Zhu, J. Dooley, and T. J. Brazil, "Simplified Volterra series based behavioral modeling of RF power amplifiers using deviation-reduction," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2006, WEPG-03.
[13] N. Le Gallou *et al.*, "Analysis of low frequency memory and influence on solid state HPA intermodulation characteristics," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2001, vol. 2, pp. 979–982.
[14] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*, 2nd ed. Norwell, MA: Kluwer, 2000.
[15] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Reading, MA: Addison-Wesley, 1995.
[16] "Connected simulation and test solutions using the Advanced Design System" Agilent Technol., Palo Alto, CA, Applicat. Notes 1394, 2000.
[17] M. S. Muha, C. J. Clark, A. A. Moulthrop, and C. P. Silva, "Validation of power amplifier nonlinear block models," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 1999, vol. 2, pp. 759–762.

Anding Zhu (S'00–M'04) received the B.E. degree in telecommunication engineering from North China Electric Power University, Baoding, China, in 1997, the M.E. degree in computer applications from Beijing University of Posts and Telecommunications, Beijing, China, in 2000, and the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2004.

He is currently a Lecturer with the School of Electrical, Electronic and Mechanical Engineering, UCD. His research interests include high-frequency nonlinear system modeling and device characterization techniques with a particular emphasis on Volterra-series-based behavioral modeling for RF PAs. He is also interested in wireless and RF system design, digital signal processing, and nonlinear system identification algorithms.

José C. Pedro (S'90–M'95–SM'99) was born in Espinho, Portugal, in 1962. He received the Diploma and Doctoral degrees in electronics and telecommunications engineering from the Universidade de Aveiro, Aveiro, Portugal, in 1985 and 1993, respectively.

From 1985 to 1993, he was an Assistant Lecturer with the Universidade de Aveiro, and a Professor since 1993. He is currently a Senior Research Scientist with the Instituto de Telecomunicações, Universidade de Aveiro, as well as a Full Professor. He has authored or coauthored several papers appearing in international journals and symposia. He coauthored *Intermodulation Distortion in Microwave and Wireless Circuits* (Artech House, 2003). His main scientific interests include active device modeling and the analysis and design of various nonlinear microwave and opto-electronics circuits, in particular, the design of highly linear multicarrier PAs and mixers.

Dr. Pedro is an associate editor for the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES and is also a reviewer for the IEEE Microwave Theory and Techniques Society (IEEE MTT-S) International Microwave Symposium (IMS). He was the recipient of the 1993 Marconi Young Scientist Award and the 2000 Institution of Electrical Engineers (IEE) Measurement Prize.

Thomas J. Brazil (M'86–SM'02–F'04) was born in County Offaly, Ireland. He received the B.E. degree in electrical engineering from University College Dublin (UCD), Dublin, Ireland, in 1973 and the Ph.D. degree from the National University of Ireland, Dublin, Ireland, in 1977.

He subsequently worked on microwave subsystem development with Plessey Research (Caswell), U.K., from 1977 to 1979. After a year as a Lecturer with the Department of Electronic Engineering, University of Birmingham, Birmingham, U.K., he returned to UCD in 1980, where he is currently a Professor with the School of Electrical, Electronic and Mechanical Engineering and holds the Chair of Electronic Engineering. He has worked in several areas of science policy, both nationally and on behalf of the European Union. From 1996 to 1998, he was Coordinator of the European EDGE project, which was the major European Union (EU) Framework *IV* (ESPRIT) project in the area of high-frequency CAD. His research interests are in the fields of nonlinear modeling and device characterization techniques with particular emphasis on applications to microwave transistor devices such as GaAs field-effect transistors (FETs), high electron-mobility transistors (HEMTs), bipolar junction transistors (BJTs), and HBTs. He also has interests in convolution-based CAD simulation techniques and microwave subsystem design.

Prof. Brazil is a Fellow of Engineers Ireland. He is a member of the Royal Irish Academy. From 1998 to 2001, he was an IEEE Microwave Theory and Techniques Society (MTT-S) worldwide Distinguished Lecturer in high-frequency applied to wireless systems. He is currently a member of the IEEE MTT-1 Technical Committee on CAD.