Pruning the Volterra Series for Behavioral Modeling of Power Amplifiers Using Physical Knowledge

Anding Zhu, *Member, IEEE*, José Carlos Pedro, *Fellow, IEEE*, and Telmo Reis Cunha, *Member, IEEE*

***Abstract*—This paper presents an efficient and effective approach to pruning the Volterra series for behavioral modeling of RF and microwave power amplifiers. Rather than adopting a pure "black-box" approach, this model pruning technique is derived from a physically meaningful block model, which has a clear linkage to the underlying physical behavior of the device. This allows all essential physical properties of the PA to be retained, but significantly reduces model complexity by removing unnecessary coefficients from the general Volterra series. A reduced-order model of this kind can be easily extracted from standard time/frequency-domain measurements or simulations, and may be simply implemented in system-level simulators. A complete physical analysis and a systematic derivation are presented, together with both computer simulations and experimental validations.**

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

I. Introduction

BEHAVIORAL modeling for RF and microwave power amplifiers (PAs) has received much attention from many researchers in recent years. It provides a convenient and efficient way to predict system-level performance without the computational complexity of full simulation or the physical analysis of nonlinear circuits, thereby significantly speeding up system design and verification process. As wireless communication is evolving towards broadband services, we increasingly encounter frequency-dependent behavior, i.e., memory effects, in RF PAs. To accurately model a PA, we have to take into account both nonlinearities and memory effects.

The Volterra series is a multidimensional combination of a linear convolution and a nonlinear power series [1]. It provides a general way to model a nonlinear dynamic system so that it can be employed to characterize a nonlinear PA with memory effects. However, since all nonlinearities and memory effects are treated equally in the classical Volterra model, the number of coefficients to be estimated increases exponentially with the degree of nonlinearity and memory length of the system. Therefore, it has been very difficult to find a practically convenient procedure for extracting full Volterra kernels of order greater than five, which restricts the practical use of the general Volterra model to the characterization of relatively weakly nonlinear PAs.

To overcome the modeling complexity, various model-order reduction approaches have been proposed to simplify the Volterra model structure. For example, in the Wiener- or Hammerstein-like models [2]–[4], memory effects are represented by linear filters, while nonlinearity is characterized by static/memoryless polynomials in a cascade arrangement. However, in a Wiener system, the pth-order Volterra kernel must be proportional to the p-folded product of their linear elements; while a Hammerstein model requires that the Volterra kernels are only nonzero along their diagonals and each kernel diagonal is proportional to the impulse response of the linear subsystem. All off-diagonal coefficients are set to zero in a *memory polynomial model* [5], while *near-diagonality reduction*-based models [6] only keep the coefficients on and near the main diagonal lines. *Polyspectral models* [7] are again based on filter/static-nonlinearity cascades, where the multidimensional nonlinear filters are approximated by 1-D versions. In the *modified/dynamic Volterra series* [8]–[11], high-order dynamics are normally omitted since they are considered to have little effect on the output of a PA. Orthonormal basis functions, like the Laguerre [12] and Kautz [13] functions, were employed as the basis for the Volterra expansion to efficiently model long-term memory effects. However, it was found difficult to locate the pre-decided poles.

Although these simplified models have been employed to characterize PAs with reasonable accuracy under certain conditions, there is no systematic way to verify if the model structure chosen is truly appropriate to the PA under study. Indeed, because behavioral models developed to date have been mainly based on a pure "black-box" approach, or were mostly constructed from "blind" nonlinear system identification algorithms (where the amplifier was considered to be a complete, or very general nonlinear system), we cannot guarantee that the relevant conditions are satisfied when doing a specific model truncation. In particular, little or no PA physical knowledge was taken into account during the model development or model-order truncation.

In this paper, we seek to construct a behavioral model for RF PAs from a physical, rather than a pure "black-box" perspective, so that we may have a clear idea on how to select a proper model

Manuscript received August 9, 2006; revised December 22, 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 Sixth Framework Program funded by the European Commission, and in part by the Portuguese Science Foundation under the ModEx Project.

A. Zhu is with the School of Electrical, Electronic and Mechanical Engineering, University College Dublin, Dublin 4, Ireland (e-mail: anding.zhu@ucd.ie).

J. C. Pedro and T. R. Cunha are with the Institute of Telecommunications, University of Aveiro, 3810-193 Aveiro, Portugal (e-mail: jcpedro@det.ua.pt; trcunha@det.ua.pt).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMTT.2007.895155

structure for a specific PA, and have an insight on how to prune it in a physically meaningful way. To achieve this, we first investigate the physical properties of a broad range of real amplifiers, i.e., the origins of their nonlinearities and short/long-term memory mechanisms. These physical behaviors are then summarized and abstracted to form a functional block model, which is sufficiently simple, but includes all essential characteristics of the PA. Since this model is not as "general" as the complete "black box" normally used in general nonlinear system identification, it follows that it becomes a *special* case of the Volterra series from which we are able to find the coefficients those are either redundant or unrelated to the actual PA physical characteristics and, thus, can be removed. This provides us a new, efficient, and effective way to prune the general Volterra series for PA behavioral modeling.

Rather than following the trial-and-error procedures used in previous modeling techniques, this model pruning strategy is directly linked to the physical behavior of the device. It thus allows us to significantly simplify the model structure and, therefore, dramatically reduces model complexity while guaranteeing that all essential physical properties of the PA are still captured. A reduced-order model of this kind has a much smaller number of coefficients, while it still has the same properties as the classical Volterra series, e.g., linearity in model parameters. Hence, it can be easily extracted from standard time/frequency-domain measurements or simulations, and simply implemented in system-level simulators.

This paper is organized as follows. In Section II, we discuss nonlinear behavior and memory effects mechanisms in a real PA, and then present a simplified block model for the PA. Based on this block model, a new model pruning approach is proposed in Section III. Model validation through both computer simulations and experimental tests is given in Section IV, with a conclusion presented in Section V.

II. PA Representation

In a wireless system, the distortion induced by a PA can be considered to arise from various origins such as voltage-dependent current sources, which are known as the device *I/V* characteristics, and nonlinear capacitances usually modeled as voltage-dependent charge sources, i.e., the device's *Q/V* characteristics. Due to the very high ratio between the operating frequency and the information bandwidth, these intrinsic nonlinearities of the device are normally treated as memoryless, or only capable of generating short-term memory effects. However, beyond these fast dynamics, the device and the circuits in which it is embedded can also generate much longer memory effects. In the first case, we have the so-called low-frequency dispersion, which includes both electrothermal nonlinear dynamics and charge carrier trapping effects. In the second case, we have the bias networks, which can involve very long time constants, and also resonances of the input and output matching networks, i.e., lightly damped impulse response tails. Fortunately, except in very wideband systems, under normal operation, the frequency of the information signal delivered by wireless PAs is much lower than the carrier frequency, and its bandwidth occupies only a negligible fraction of the PA available

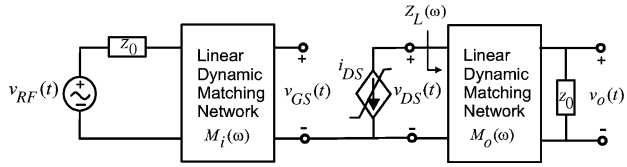

Fig. 1. Simplified circuit schematic of an FET-based PA.

bandwidth so that the matching networks can be considered almost flat. In other words, the device's input and output terminating impedances are memoryless to slowly varying complex envelopes, except where their bias networks are concerned. In fact, if the PA suffers any bias variations determined by the input amplitude modulation, the dc supply voltage will then vary according to the slow dynamics of the bias networks.

To understand this process, we can start by the simplified schematic model of a single-stage PA shown in Fig. 1. In this circuit schematic, the active device, in this case, a field-effect transistor (FET), was assumed as showing no internal feedback, i.e., negligible gate–drain capacitance $C_{\rm GD}$ or source resistance and inductance R_S and L_S, and its input and output linear resistance and capacitance components were lumped into the input and output matching networks. The nonlinear active device is thus represented by its nonlinear output current source $i_{\rm DS}$, which is dependent on the input $v_{\rm GS}$, and the output $v_{\rm DS}$, control voltages, i.e., $i_{\rm DS}(v_{\rm GS}, v_{\rm DS})$. The input control signal voltage $v_{\rm GS}(t)$ is simply a linearly filtered replica of the input excitation $v_{\rm RF}(t)$, but the determination of the output control signal voltage $v_{\rm DS}(t)$ is much more complex because of the $i_{\rm DS}(v_{\rm GS}, v_{\rm DS})$ nonlinearity and its interaction with the output matching and bias networks.

In fact, if we use $Z_L(\omega)$ to represent the impedance shown by those matching and bias components to the $i_{\rm DS}$ current source, we obtain the following equations for the PA operation in the frequency domain:

$$V_{\rm GS}(\omega) = M_i(\omega) V_{\rm RF}(\omega) \tag{1}$$

$$I_{\rm DS}(\omega) = \mathbb{F}\left\{ i_{\rm DS}\left[v_{\rm GS}(t), v_{\rm DS}(t) \right] \right\} \tag{2}$$

$$V_{\rm DS}(\omega) = Z_L(\omega) I_{\rm DS}(\omega) \tag{3}$$

where $\mathbb{F}\{\cdot\}$ denotes the conventional time-to-frequency Fourier transform. Although the model of Fig. 1 seems to be a cascade model, the interaction between the static nonlinearity $i_{\rm DS}(v_{\rm GS}, v_{\rm DS})$ and the output dynamic linear filter can be viewed as a feedback process. Indeed, while the dependence of $V_{\rm GS}(\omega)$ on $V_{\rm RF}(\omega)$ and of $I_{\rm DS}(\omega)$ on $V_{\rm GS}(\omega)$ can be described by a linear and nonlinear transfer function, respectively, the dependence of $I_{\rm DS}(\omega)$ on $V_{\rm DS}(\omega)$ involves the following feedback process. Due to its nonlinear dependence on $V_{\rm GS}(\omega)$, $I_{\rm DS}(\omega)$ incorporates linear and nonlinear frequency components involving all types of fundamental, harmonic, and baseband mixing products. Flowing through the output impedance $Z_L(\omega)$, these mixing products will be converted into voltage components with both short- and long-term memory, just as if the current variable $I_{\rm DS}(\omega)$ flowed through a linear filter of transfer function $Z_L(\omega)$ producing a voltage output $V_{\rm DS}(\omega)$. This voltage output is then nonlinearly remixed back with the original drain–source current because $i_{\rm DS}$ also depends on $v_{\rm DS}$.

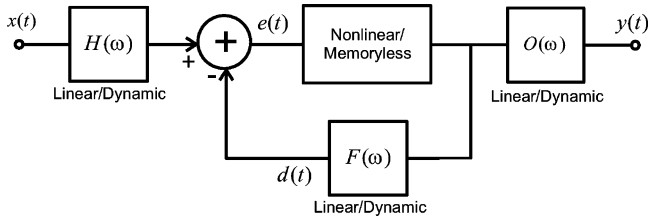

Fig. 2. Conceptual feedback model of the PA.

In conclusion, as was first explained in detail in [14], and then followed by other researchers [15], these nonlinearity-memory interactions in the PA can be modeled by a conceptual feedback block model shown in Fig. 2. It uses a general static nonlinearity, as the feedforward path, to represent the nonlinear transformation of $i_{\rm DS}(v_{\rm GS}, v_{\rm DS})$, and a linear filter in the feedback loop to represent the action of the dynamic output impedance $Z_L(\omega)$. This emulates the interactions between the PA's memoryless nonlinearities and the memory effects imposed by the linear dynamic circuitry in which they are embedded, even if this network is simply an equivalent circuit, as is the case of the electro-thermal dynamics. Beyond the core nonlinearity and the dynamic feedback loop, the functional block diagram of Fig. 2 also includes one input and one output filter $H(\omega)$ and $O(\omega)$, which represents the input and output matching networks of the PA $M_i(\omega)$ and $M_o(\omega)$, respectively.

Since this block model is only a conceptual view, it may not be amenable for direct extraction from practical measurement data sets. However, as discussed in [14], the most important advantage of this feedback structure is that it is sufficiently simple to allow a rigorous Volterra series analysis, while still keeping the PA's essential nonlinear dynamic characteristics. Furthermore, from this model, we can see that, although a PA is a nonlinear dynamic system showing a very complex nonlinear dynamic behavior, it is not as "general" as a pure "black-box" and, therefore, it can be considered as a particular case of the general Volterra series. Hence, it should be possible to prune the Volterra series, retaining only the specific coefficients' subsets that are necessary for representing the referred feedback block, but deleting all other ones, as proposed in the following.

III. Pruning the Volterra Model

In the discrete time domain, a Volterra series can be written as

$$y(n) = \sum_{p=1}^{P} y_p(n) \tag{4}$$

where $y_p(n)$ represents the contribution of the pth-order nonlinearity, and

$$y_p(n) = \sum_{i_1=0}^{M} \cdots \sum_{i_p=0}^{M} h_p(i_1, \ldots, i_p) \prod_{j=1}^{p} x(n - i_j) \tag{5}$$

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, as is assumed in (4) and (5), the Volterra series is normally truncated to finite nonlinear order P and finite memory length M [1]. To derive a Volterra model for the PA in Fig. 2, a common approach is the harmonic probing method, usually conducted in the frequency domain [16]. That method is straightforward for the first few nonlinear orders, but it quickly becomes cumbersome when high-order nonlinearities are involved. In this paper, we directly derive the Volterra model in the discrete time domain. Before proceeding, however, we first make several simplifications and assumptions for the block model in Fig. 2.

The first simplification is that we remove the two linear filter blocks $H(\omega)$ and $O(\omega)$. This is reasonable because these filters stand for the input and output matching networks, which, under the PA's normal operation, and as explained in Section II, behave in a memoryless way to the slowly varying complex envelopes in which we are interested.

Second, it is assumed that, although the model of Fig. 2 is a system with infinite memory due to its dynamic feedback path, it can still be represented by a feedforward finite memory system such as a truncated Volterra series. This can be justified for at least two reasons. Firstly, from a physical point-of-view, it is obvious that the PA output does not depend on the input's infinitely remote past. Second, it is known that the result of the convolution of the feedback linear dynamic filter impulse response with the excitation has a time duration that is longer than the one of the original excitation (it is, in fact, the sum of the length of the excitation and the length of the filter impulse response), similar to the way in which the feedforward nonlinearity creates spectral widening from its input excitations due to the convolution of spectra. Hence, to guarantee that the feedback system can, in fact, be modeled with finite memory, we need to truncate the system's output memory span, as we would truncate the frequency domain output harmonic content of the nonlinearity. For that, we first assume that the memory span of the overall system can be truncated to M, in which all necessary previous input information is taken into account. Second, we consider that the impulse response of the feedback filter $F(\omega)$ has that same memory span, even if, for that purpose, some of its coefficients are set to zero after its own natural memory span M_F (assuming $M_F < M$). In this sense, we can conclude that, in the discrete time domain, to truncate the feedback loop to an approximated feedforward system, we could assume that the components at the output of the nonlinear block only enter the filter once since the second or following entries would be out of the system's memory span. From a physical point-of-view, this memory span truncation is reasonable since the items after second entries would either be mixed up to generate higher order components or become far away from the current input, producing an impact on the current output that should be negligible. Moreover, it is also consistent with the cascaded nonlinearity–linear filter-nonlinearity structure presented in [17] and [18], which, as discussed in [2], can be understood as an unfolded, or feedforward, version of the feedback structure of Fig. 2. This leads to the conclusion that, in the discrete time domain, all output items with delays, e.g., $x^3(n-i)$, or products with delayed terms, e.g., $x(n)x^2(n-i)$, will not enter the filter again since they (or part of them) have already passed through the feedback loop so that only items without any delays, such as $x^2(n), x^3(n), \ldots$, will enter the filter and be fed back to the input.

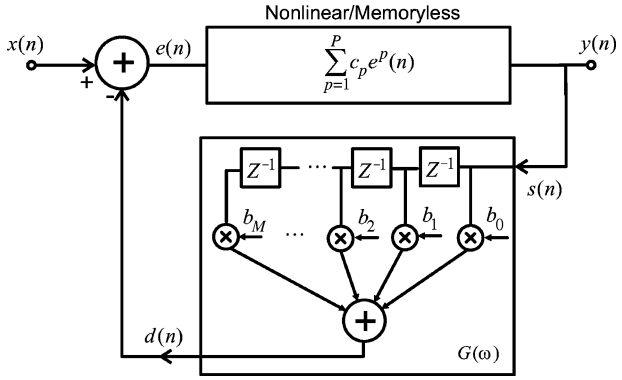

Fig. 3. Equivalent PA block model in the discrete time domain.

The third assumption we make on the model of Fig. 2 is that the feedback filter $F(\omega)$ is flat at the fundamental frequency band because the bandwidth of the PA excitation $X(\omega)$ is assumed to be narrow compared to the linear system's frequency response $S_1(\omega)$. That must be true because, as explained above, in typical wireless systems, the relative excitation bandwidth BW/f_0 is very small, and it is much smaller than the one imposed by the PA filters' quality factor Q. Since $S_1(\omega)$ can be considered flat, and it is related to $F(\omega)$ by $S_1(\omega) = a_1/[1 - a_1F(\omega)]$, where a_1 is a constant [14], this implies that $F(\omega)$ must also be flat at the fundamental frequency band. This results in $F(\omega)$ behaving as a memoryless block to any components of the output $y(n)$, whose frequency falls in the system's fundamental frequency band. In other words, for that first zone output, the frequency domain coefficients of $F(\omega)$ are a constant and its time domain impulse response is a unique Dirac delta function. Therefore, we can separate this fundamental frequency response—a mere scalar operation—from the remaining frequency bands of the filter $F(\omega)$, and merge it into the memoryless block. The new static nonlinearity block can then be represented by a Pth-order polynomial function, while the rest of the characteristics of $F(\omega)$ are used to form a new filter $G(\omega)$, whose impulse response to the fundamental frequency is zero. In the discrete time domain, $G(\omega)$ can be represented by a transversal finite impulse response (FIR) filter with memory length M.

In summary, the block model of Fig. 2 can be transferred to the equivalent model in the discrete time domain, as shown in Fig. 3, from which we now develop an equivalent Volterra series representation.

As discussed earlier, the impulse response of the feedback filter $G(\omega)$ to the fundamental frequency is zero, which means that the original input signal $x(n)$ will not enter the filter at the output, and considering the system has finite memory and its memory span is equal to the memory length of the feedback filter, the delayed terms at the output will not enter the filter again. This has the consequence that the input signal of the feedback filter $s(n)$ will include only terms that are nonlinear and without any delays, such as $x^2(n), x^3(n), \ldots$, i.e.,

$$s(n) = \sum_{p=2}^{P} c_p x^p(n) \tag{6}$$

where c_p represents the scalar factor of $x^p(n)$.

When $s(n)$ passes the feedback loop, the filter $G(\omega)$ will create tails to these nonlinear terms. For example, for the second-order term $x^2(n)$, the output $d[s(n)]$ will be

$$d\left[x^2(n)\right] = c_2 \sum_{i=0}^{M} b_i x^2(n-i) \tag{7}$$

where b_i is the coefficient of the filter $G(\omega)$. These tails will be remixed with the original RF signal to create nonlinear distortions and memory effects. This happens to other high-order terms in the same way.

From (6) and (7), we conclude that the output of the filter $d(n)$ can be formulated as

$$d(n) = \sum_{p=2}^{P} c_p \sum_{i=0}^{M} b_i x^p(n-i) \tag{8}$$

which can be considered as a linear combination of $x^2(n - i), x^3(n - i), \ldots$. The error signal $e(n)$ then becomes

$$e(n) = x(n) - \sum_{p=2}^{P} c_p \sum_{i=0}^{M} b_i x^p(n-i) \tag{9}$$

which is also a linear function of $x^2(n - i), x^3(n - i), \ldots$, plus $x(n)$. Finally, when $e(n)$ passes the memoryless block in the feedforward path, the polynomial function becomes a series of multinomial operations to the individual input items $x(n), x^2(n-i), x^3(n-i), \ldots$, in which these items are mixed together to generate the whole set of PA nonlinear distortions and memory effects. For instance, the contributions to the third-order distortion will come from: 1) three $x(n)$ samples mixed together by the third degree polynomial term $c_3e^3(n)$ and 2) one $x(n)$ mixed with one $x^2(n-i)$ by the second degree polynomial term $c_2e^2(n)$. Note that only remixing components are taken into account here. The components that are arising directly from the first degree polynomial term $c_1e(n)$, such as $x^3(n-i)$ in this case, are omitted. This is because the fundamental parts generated from these terms are zero when they pass the feedback filter since $G(\omega)$ is zero at the fundamental frequency band so that they do not affect the output $y(n)$ in the first zone. The higher order distortions can be derived in the same way.

In conclusion, the output $y(n)$ will be a sum of product terms of the multinomial functions. The coefficients, corresponding to these items, will be products of the coefficients of the polynomial function c_p, and the coefficients of the feedback filter b_i, scaled by the indices of the multinomial functions. These coefficients cannot be easily identified directly since products are involved. However, they can be regrouped and generalized to form equivalent Volterra kernels in the classical Volterra format. For example, $c_3b_1b_2$ can be transferred to $h_5(0, 1, 1, 2, 2)$, which corresponds to the input item $x(n)x^2(n-1)x^2(n-2)$. Some samples of these Volterra kernels and their corresponding input items are listed in Table I. From that table, we can immediately derive the contributions for different order nonlinearities as follows.

- First order

$$y_1(n) = h_1(0)x(n). \tag{10}$$

TABLE I
INPUT ITEMS AND THEIR CORRESPONDING COEFFICIENTS

	1st-order		3rd-order		5th-order		...
	Input Items	Corresponding Coefficients	Input Items	Corresponding Coefficients	Input Items	Corresponding Coefficients	...
$e(n)$	$x(n)$	$h_1(0)$					
$e^2(n)$			$x(n)x^2(n\text{-}i)$	$h_3(0,i,i)$	$x^2(n\text{-}i_1)x^3(n\text{-}i_2)$ $x(n)x^4(n\text{-}i)$	$h_5(i_1,i_1,i_2,i_2,i_2)$ $h_5(0,i,i,i,i)$	
$e^3(n)$			$x(n)x(n)x(n)$	$h_3(0,0,0)$	$x(n)x^2(n\text{-}i_1)x^2(n\text{-}i_2)$ $x(n)x(n)x^3(n\text{-}i)$	$h_5(0,i_1,i_1,i_2,i_2)$ $h_5(0,0,i,i,i)$	...
$e^4(n)$					$x(n)x(n)x(n)x^2(n\text{-}i)$	$h_5(0,0,0,i,i)$	
$e^5(n)$					$x(n)x(n)x(n)x(n)x(n)$	$h_5(0,0,0,0,0)$	
...							...

- Third order

$$y_3(n) = h_3(0,0,0)x^3(n) + \sum_{i=1}^{M} h_3(0,i,i)x(n)x^2(n-i). \quad (11)$$

- Fifth order

$$\begin{aligned} y_5(n) = {} & h_5(0,0,0,0,0)x^5(n) \\ & + \sum_{i=1}^{M} h_5(0,0,0,i,i)x^3(n)x^2(n-i) \\ & + \sum_{i=1}^{M} h_5(0,0,i,i,i)x^2(n)x^3(n-i) \\ & + \sum_{i_1=1}^{M} \sum_{i_2=i_1+1}^{M} h_5(0,i_1,i_1,i_2,i_2) \\ & \times x(n)x^2(n-i_1)x^2(n-i_2) \\ & + \sum_{i=1}^{M} h_5(0,i,i,i,i)x(n)x^4(n-i) \\ & + \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} h_5(i_1,i_1,i_2,i_2,i_2) \\ & \times x^2(n-i_1)x^3(n-i_2) \end{aligned} \quad (12)$$

and so on.

Compared to (5), we can see that now the general multidimensional convolutions are reduced to 1-D or 2-D ones so that only a small subset of Volterra kernels appears in (10)–(12). The remaining coefficients are considered to be either zero and unrelated to the PA output behavior or merged into the coefficients on the list, which are redundant with the ones already present in (10)–(12). Hence, the total number of coefficients increases only almost linearly with the nonlinearity order or memory length. This significantly reduces the modeling complexity. For example, in the full Volterra model, a fifth-order expansion with memory length 8 would lead to a total number of coefficients of 59 049 or 1287, considering symmetry, while the new pruned model only involves 117 parameters.

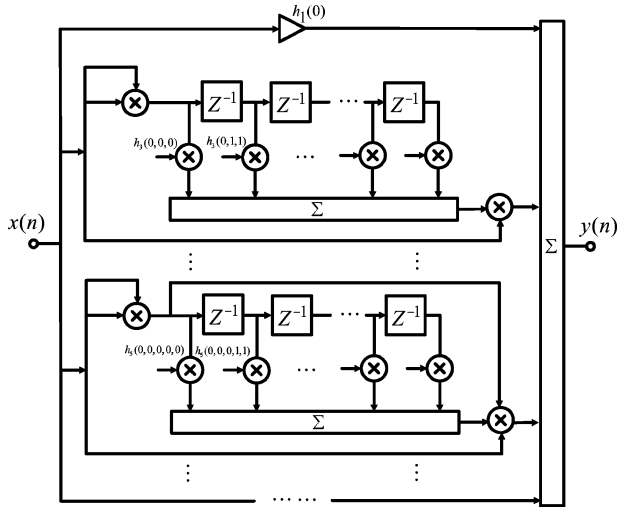

Fig. 4. Sample of the pruned Volterra model implementation.

While the reduced-order model has much smaller number of coefficients, it still has the same properties as in the classical Volterra series, e.g., the output of the model is also linear with respect to the coefficients, so that it can be extracted directly by employing linear estimation algorithms in the discrete time domain. Furthermore, because the number of coefficients is dramatically reduced, the model extraction becomes much easier. Model implementation is also significantly simplified since only a limited number of multiplier products and convolutions are needed, as shown in Fig. 4. This model can be systematically extended to higher orders without any further difficulties because its input items are simple products from multinomial functions, as shown in Table I.

Finally, note that, in the derivation above, only real RF signals were considered. For handling complex envelope signals, these Volterra coefficients have to be transformed to a low-pass equivalent format, as is explained in the Appendix.

IV. MODEL VALIDATION

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

A. Computer Simulations

In this first test, we designed an equivalent-circuit PA model and simulated it with the Agilent's Advanced Design System (ADS) [19] simulation software package. This is a GaAs MESFET class-A PA operating at 2 GHz, excited by 3GPP W-CDMA signals of 3.84-Mc/s chip rate. The reason for using computer simulations was that this virtual test setup enabled us to easily control the PA nonlinearity and memory effects, and also allowed us to eliminate noise and measurement errors, which may mask the actual model accuracy.

This PA was simulated by a co-simulation of Ptolemy and the Circuit-Envelope Simulator in ADS 2004A [19]. Although the proposed model can be employed to represent a wide range of the PA's nonlinear characteristics and memory effects, as the general Volterra model, in this test, we only concentrated on memory effects arising from the bias networks. Other memory effects, such as self-heating and trapping effects, were not considered since the MESFET nonlinear model did not include them.

To investigate the capability of our model in representing PA memory effects, we simulated the amplifier circuit under two different bias networks, which were: 1) ideal, in which the dc feed is close to the ideal short circuit and 2) nonideal, in which the dc feed shows a nonnegligible impedance to the envelope frequency components. The resulting dynamic AM/AM plots are shown in Fig. 5. From these plots, we can see that the PA did not present any significant memory under ideal bias networks, while memory became evident (AM–AM plots showing distinct hysteresis loops) when the bias impedance increased, something to be expected from a real PA. As discussed in Section II, these memory effects were mainly present in the nonlinear operating region since they arise from remixing the original input with low/high-frequency harmonics and intermodulation products fed back from the output.

Fifty sets of time-domain envelope waveforms were captured from the input and output of the PA under different output power levels, and with a sampling rate of 30.72 MHz. These data were then used for model extraction and model validation. The model was truncated to fifth-order nonlinearity with memory length from three to eight, and was extracted via a least squares (LS) algorithm in the discrete time domain. A sample of the output time domain complex envelopes' magnitude and phase are shown in Fig. 6(a) and (b), respectively. These results clearly show that the modeled data indeed fitted the desired outputs very well. The normalized mean square errors (NMSEs) were calculated for various validation data, and the average of them was approximately -43 dB, which indicates that the relative errors between the modeled and simulated time domain outputs were less than 0.005%. For comparison, a fifth-order complex polynomial (memoryless) model was also extracted for this PA, whose output waveforms are shown in Fig. 6. Although the phase part was fitted well, errors appeared in the magnitude. The NMSE for this model only reached -29 dB. To show the model accuracy in the frequency domain, the spectra of modeled errors are plotted in Fig. 7. There we can see that the error signal spectrum of the new model is almost close to the noise floor, while significant errors are generated in the output predicted by the memoryless model. For reference, the spectrum of the simulated output is also plotted in Fig. 7.

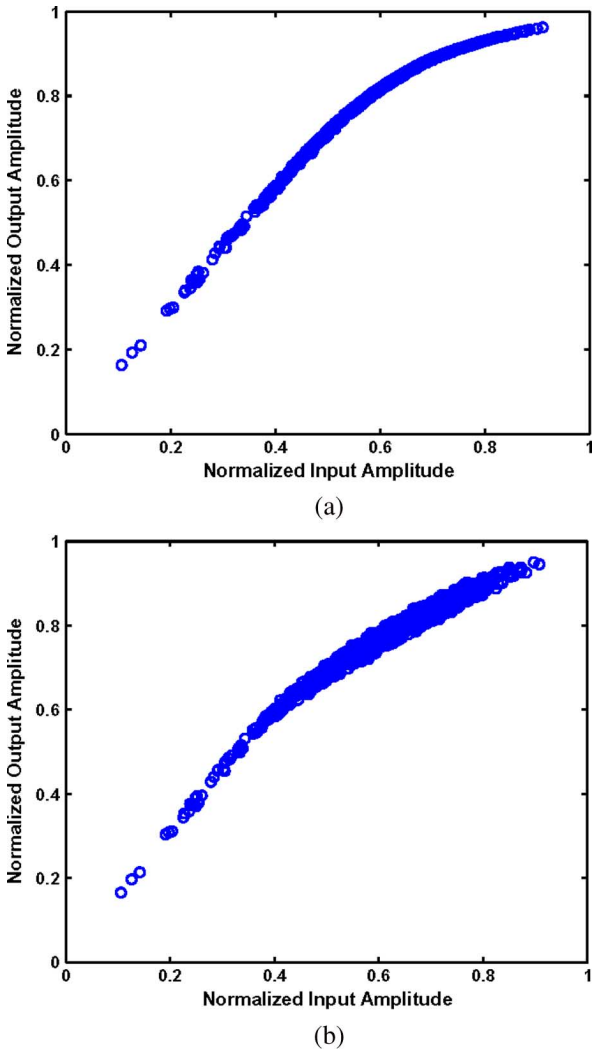

Fig. 5. Sample AM/AM plots for the PA with: (a) ideal bias networks and (b) nonideal bias networks.

B. Experimental Tests

To make this modeling technique closer to the "real" world, we also tested a commercial LDMOS class-AB PA in our laboratory. Its schematic diagram is depicted in Fig. 8. This PA, operated at 2.14 GHz, and was excited by W-CDMA signals of a 3.84-Mc/s chip rate and with 8.2-dB peak-to-average power ratio (PAPR). The average output power of the PA is 10 W, and its AM/AM characteristics were close to the ones seen in the first simulated PA circuit.

The test bench setup used the ADS–electronic signal generator (ESG)–VSA connected solution [20]. The modulated W-CDMA data files were first created at baseband, downloaded to the arbitrary waveform generator, as complex in-phase (I) and quadrature (Q) signals, and were then fed to the IQ modulator present in the ESG. This generator was used to produce the RF test signal to the PA. The output of the PA was then down-converted and sampled by the vector signal analyzer (VSA). To eliminate noise and measurement errors, 30 repeated

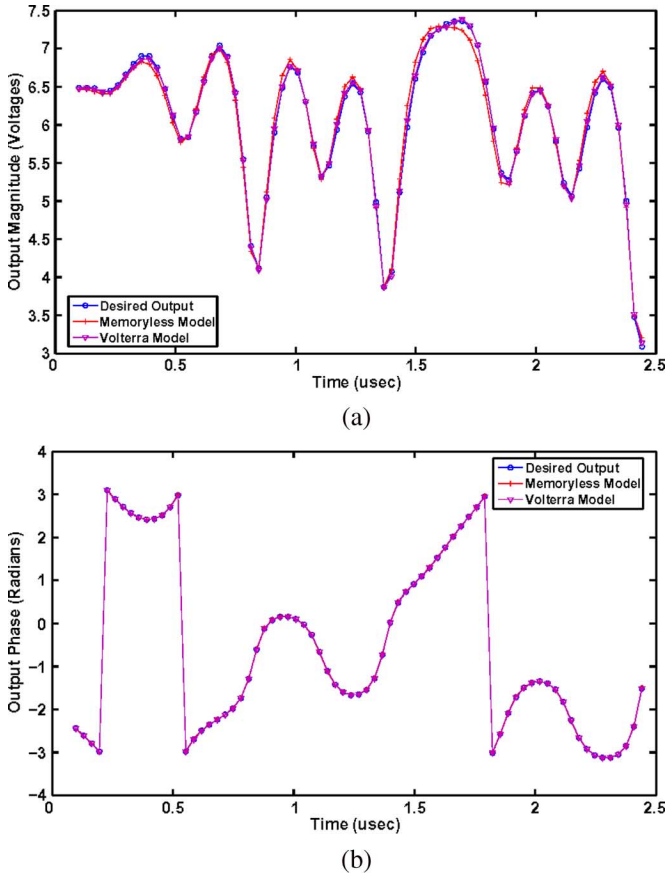

Fig. 6. Sample time domain complex envelope output waveforms of modeled and simulated: (a) magnitude and (b) phase.

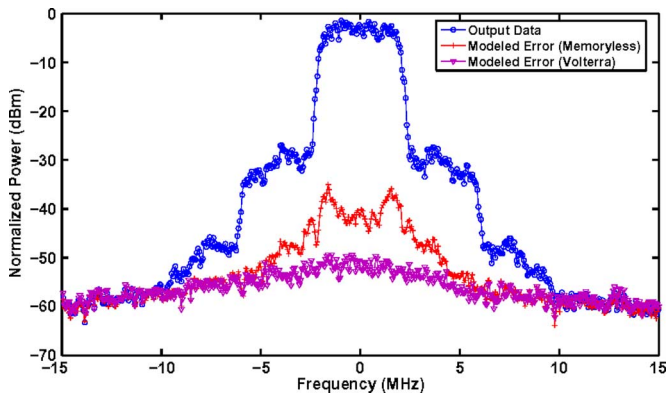

Fig. 7. Sample frequency domain output and modeled error spectra.

measurements were performed, and around 150 000 sampling data, with a sampling rate of 30 MHz, were captured from the PA input and output envelope signals. These data were pre-processed, via averaging and alignment, before they were used for model extraction and model validation. The model was extracted in the same way as in the previous verification tests via simulation.

The time domain waveforms of real and imaginary parts of the PA output complex envelopes are shown in Fig. 9(a) and (b), respectively. They indicate that the measured data points were again well fitted by the modeled ones. The average NMSE was,

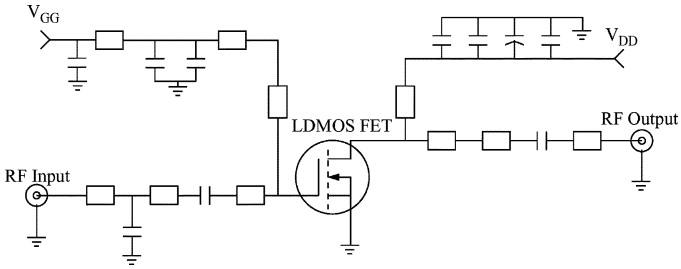

Fig. 8. Schematic diagram of the tested PA.

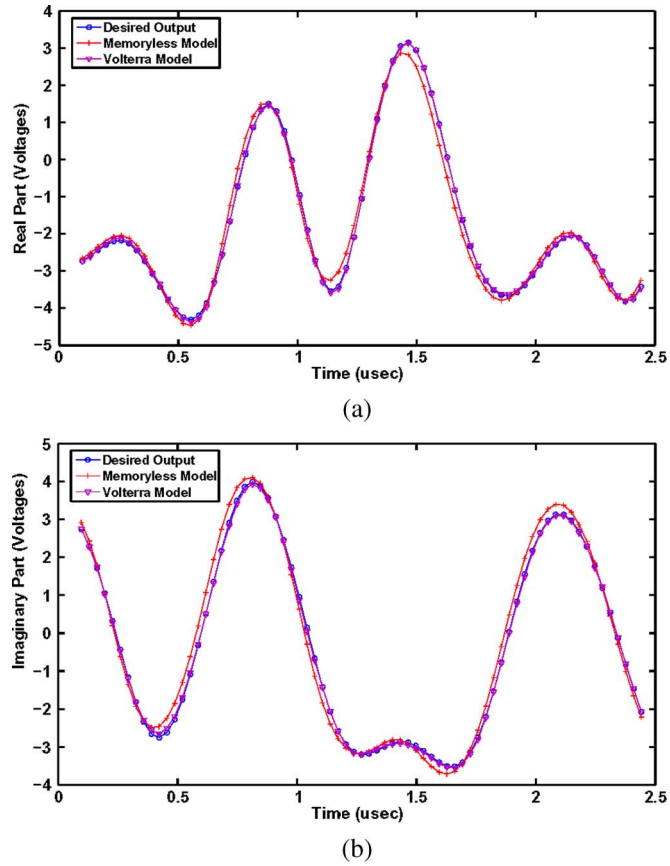

Fig. 9. Sample time domain complex envelope output waveforms of modeled and measured: (a) real part and (b) imaginary part.

in this case, −38.2 dB, which was a little higher than that of the simulation because of noise and measurement errors. The output waveforms predicted by the memoryless polynomial model are also plotted in Fig. 9, and the NMSE for that model was only −24 dB, which indicates the occurrence of significant modeling errors. The model performance when predicting PA gain and the adjacent channel power ratios (ACPRs) are shown in Table II. We can see that the measured results were accurately predicted by the proposed model.

Although in the above validation tests we only demonstrated the model working up to fifth-order nonlinearity and eight time-delay memory lengths, this model can be easily extended to higher orders and longer memory lengths. This is because, by employing the model pruning approach proposed in Section III, the number of coefficients of the model can be kept reasonably small even if higher orders and longer term memory are involved

TABLE II
GAIN AND ACPR PERFORMANCE

Performance	Gain (dB)	ACPR (dBc) (@+/- 5MHz offset)	ACPR (dBc) (@+/- 10MHz offset)
Measured	13.1	-39.2/-37.6	-54.2/-53.5
Volterra Model	13.0	-38.5/-36.8	-53.9/-53.2
Memoryless Model	12.8	-41.4/-41.4	-56.7/-56.7

since this number increases almost linearly with the order of the nonlinearity or memory length.

V. CONCLUSION

An efficient and effective Volterra model pruning method for RF PAs has been presented in this paper. The advantage of this model reduction approach is that it allows efficient reduction of the model complexity, while keeping all essential physical properties of a real PA since it was derived from a functional block model, which has a clear linkage to the device's physical behavior. Both computer simulation and experimental verification tests indicated that this model can be employed to model a PA with very high accuracy, but with a much smaller number of coefficients than the commonly used general Volterra models.

APPENDIX

In system level analysis and design, most simulators use baseband complex envelope signals to evaluate the system performance since modulation techniques are normally employed to carry useful information. For handling these carrier-modulated signals, the real bandpass Volterra coefficients and their corresponding inputs have to be transformed into the complex envelope format. For example, the real kernel $h_3(i_1,i_2,i_3)$ becomes the complex kernel $\tilde{h}_3(i_1,i_2,i_3^*)$ where i_3^* indicates a complex conjugate transform need be made to its corresponding input term $\tilde{x}(n-i_3)$, namely, its corresponding input is $\tilde{x}(n-i_1)\tilde{x}(n-i_2)\tilde{x}(n-i_3)^*$, where $(\cdot)^*$ represents the complex conjugate transform. The details of the transforms are as follows.

- First order

$$h_1(0) \Rightarrow \tilde{h}_1(0). \tag{a1}$$

- Third order

$$h_3(0,0,0) \Rightarrow \tilde{h}_3(0,0,0^*) \tag{a2}$$

$$h_3(0,i,i) \Rightarrow \begin{cases} \tilde{h}_3(0,i,i^*) \\ \tilde{h}_3(0^*,i,i). \end{cases} \tag{a3}$$

- Fifth order

$$h_5(0,0,0,0,0) \Rightarrow \tilde{h}_5(0,0,0,0^*,0^*) \tag{a4}$$

$$h_5(0,0,0,i,i) \Rightarrow \begin{cases} \tilde{h}_5(0,0,0,i^*,i^*) \\ \tilde{h}_5(0,0,0^*,i,i^*) \\ \tilde{h}_5(0,0^*,0^*,i,i) \end{cases} \tag{a5}$$

$$h_5(0,0,i,i,i) \Rightarrow \begin{cases} \tilde{h}_5(0,0,i,i^*,i^*) \\ \tilde{h}_5(0,0^*,i,i,i^*) \\ \tilde{h}_5(0^*,0^*,i,i,i) \end{cases} \tag{a6}$$

$$h_5(0,i_1,i_1,i_2,i_2) \Rightarrow \begin{cases} \tilde{h}_5(0,i_1,i_1,i_2^*,i_2^*) \\ \tilde{h}_5(0,i_1,i_1^*,i_2,i_2^*) \\ \tilde{h}_5(0^*,i_1,i_1^*,i_2,i_2) \end{cases} \tag{a7}$$

$$h_5(0,i,i,i,i) \Rightarrow \begin{cases} \tilde{h}_5(0,i,i,i^*,i^*) \\ \tilde{h}_5(0^*,i^*,i,i,i) \end{cases} \tag{a8}$$

$$h_5(i_1,i_1,i_2,i_2,i_2) \Rightarrow \begin{cases} \tilde{h}_5(i_1,i_1,i_2,i_2^*,i_2^*) \\ \tilde{h}_5(i_1,i_1^*,i_2,i_2,i_2^*) \\ \tilde{h}_5(i_1^*,i_1^*,i_2,i_2,i_2). \end{cases} \tag{a9}$$

The higher order kernels can be derived in the same way.

REFERENCES

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

[2] 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.

[3] C. P. Silva *et al.*, "Optimal-filter approach for nonlinear power amplifier modeling and equalization," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Boston, MA, Jun. 2000, pp. 437–440.

[4] H. Ku, M. Mckinley, and J. S. Kenney, "Quantifying memory effects in RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 12, pp. 2843–2849, Dec. 2002.

[5] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, Nov. 2001.

[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] C. Silva, A. Moulthrop, and M. Muha, "Introduction to polyspectral modeling and compensation techniques for wideband communications systems," in *58th ARFTG Conf. Dig.*, San Diego, CA, Nov. 2001, pp. 1–15.

[8] D. Mirri *et al.*, "A nonlinear dynamic model for performance analysis of large-signal amplifiers in communication systems," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 2, pp. 341–350, Apr. 2004.

[9] E. Ngoya *et al.*, "Accurate RF and microwave system level modeling of wideband nonlinear circuits," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Boston, MA, Jun. 2000, vol. 1, pp. 79–82.

[10] 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, pp. 1113–1116.

[11] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[12] A. Zhu and T. J. Brazil, "RF power amplifiers behavioral modeling using Volterra expansion with Laguerre functions," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2005, pp. 963–966.

[13] M. Isaksson and D. Rönnow, "A Kautz–Volterra behavioral model for RF power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2006, pp. 485–488.

[14] J. C. Pedro, N. B. Carvalho, and P. M. Lavrador, "Modeling nonlinear behavior of bandpass memoryless and dynamic systems," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Philadelphia, PA, Jun. 2003, vol. 3, pp. 2133–2136.

[15] E. Ngoya and A. Soury, "Envelope domain methods for behavioral modeling," in *Fundamentals of Nonlinear Behavioral Modeling for RF and Microwave Design*. Norwood, MA: Artech House, 2005, ch. 3, pp. 37–86.

[16] J. C. Pedro and N. B. Carvalho, *Intermodulation in Microwave and Wireless Circuits*. Norwood, MA: Artech House, 2003.

[17] J. Vuolevi, T. Rahkonen, and J. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 8, pp. 1383–1389, Aug. 2001.

[18] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*. Norwood, MA: Artech House, 2003.

[19] Advanced Design System (ADS) 2004A. Agilent Technol., Palo Alto, CA [Online]. Available: http://eesof.tm.agilent.com/

[20] "Connected simulation and test solutions using the advanced design system," Agilent Technol., Palo Alto, CA, Applicat. Notes 1394, 2000.

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 Carlos Pedro (S'90–M'95–SM'99–F'07) 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 coauthored *Intermodulation Distortion in Microwave and Wireless Circuits* (Artech House, 2003) and has authored or coauthored several papers appearing in international journals and symposia. 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 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.

Telmo Reis Cunha (M'05) was born in Porto, Portugal, in 1973. He received the Diploma and Doctoral degrees in electronics and computer engineering from the Universidade do Porto, Porto, Portugal, in 1996 and 2003, respectively.

From 1997 to 2001, he was with the Observatório Astronómico, Universidade do Porto, where he was involved with diverse national and international research projects in the areas of satellite navigation and system integration. From 2001 to 2004, he was a Technical Director and Research Engineer with Geonav Ltd., a private company located near Porto, Portugal. Since 2004, he has been an invited Auxiliary Professor with the Universidade de Aveiro, and also a Research Engineer with the Instituto de Telecomunicações. His current main research interests include behavioral modeling applied to RF and microwave devices.

Dr. Cunha was the recipient of the 1997 Fundação António de Almeida Prize. He was also the recipient of the 2001 Best Presentation Award for his presentation at the ION–GPS Conference, Salt Lake City, UT.