An Overview of Volterra Series Based Behavioral Modeling of RF/Microwave Power Amplifiers

Anding Zhu, and Thomas J. Brazil

RF & Microwave Research Group
University College Dublin
Dublin 4, Ireland
{anding.zhu; tom.brazil}@ucd.ie

Abstract

This paper presents an overview of recently-developed, simplified Volterra series based, behavioral modeling approaches for radio frequency and microwave power amplifiers. Various model topologies and model pruning strategies are discussed, together a presentation of their comparative advantages and limitations.

1. Introduction

Behavioral modeling of RF power amplifiers (PAs) has received much attention from many researchers in recent years [1]. In this kind of system-level modeling, the PA is considered as a "black-box", i.e., in principle, no knowledge of internal structures is required and the modeling information is completely contained in external responses, which makes model structures significantly simplified and therefore enables fast prediction of system performances.

In a wideband wireless system, the distortion induced by a power amplifier can be considered to arise from different sources or can be assigned to different physical phenomena: (I) Static (device) nonlinearities; (II) Linear memory effects, arising from time delays, or phase shifts, in the matching networks and the device/circuit elements used; and (III) Nonlinear memory effects, such as those caused by trapping effects, non-ideal bias networks, or temperature dependence on the input power, etc.. To accurately model a PA, an accurate description of high and low frequency memory effects must be integrated into inherently nonlinear characteristics, to enable an accurate prediction of system performances.

A Volterra series is a combination of linear convolution and a nonlinear power series; it provides a general way to model a nonlinear system with memory [2][3], so that it can be employed to describe the relationship between the input and the output of an amplifier with memory. However, high computational complexity makes general methods of this kind rather impractical in some real applications because the number of parameters to be estimated in the Volterra model increases exponentially with the degree of nonlinearity and with the memory length of the system.

In recent years, extensive research work has been carried out to simplify the structure of the general Volterra series based behavioral models, aimed at reducing the complexity of this kind of method for real implementations. Different model topologies or various model pruning strategies may lead to a large disparity of model capability and model accuracy. This paper will present an overview of these up-to-date techniques, discuss their model structures and modeling capabilities, and thus provide the user with a clear guideline on how to select a proper model to characterize a real power amplifier under certain conditions.

2. Volterra series

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

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

where $x(n)$ and $y(n)$ represents the input and the output, respectively, and $h_p(i_1, \cdots, i_p)$ 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 [3].

The Volterra series in (1) can be directly employed to represent the nonlinear transfer function of a power amplifier. However, in system level analysis and design, most simulators use baseband complex envelope signals to evaluate the system performance since modulation techniques are normally employed in modern wireless communication systems, where only the envelopes carry the useful information [4]. To

handle these carrier-modulated signals, the Volterra series has to be transferred to the low-pass equivalent format.

Consider $x(t) = \Re e\left[\tilde{x}(t) \cdot e^{j\omega_0 t}\right]$ and $y(t) = \Re e\left[\tilde{y}(t) \cdot e^{j\omega_0 t}\right]$ as the input and output signal of a power amplifier, where ω_0 is carrier frequency, $\tilde{x}(t)$ and $\tilde{y}(t)$ is the envelope of input and output signal, respectively. Using A/D conversion, a discrete time-domain finite-memory complex baseband Volterra model of power amplifier has the form:

$$\tilde{y}(n) = \sum_{i=0}^{M} \tilde{h}_1(i) \times \tilde{x}(n-i)$$
$$+ \sum_{i_1=0}^{M}\sum_{i_2=i_1}^{M}\sum_{i_3=0}^{M} \tilde{h}_3(i_1, i_2, i_3) \times \tilde{x}(n-i_1)\tilde{x}(n-i_2)\tilde{x}^*(n-i_3) \quad (2)$$
$$+ \cdots$$

where $\tilde{h}_p(i_1, i_2, \cdots, i_p)$ is the pth-order *complex Volterra kernel*, and $(\cdot)^*$ represents the conjugate transpose. In the above equation, we have removed the redundant items associated with kernel symmetry, and also the even-order kernels, whose effects can be omitted in band-limited modulation systems.

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, e.g., Least Squares. 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. This makes general methods of this kind impractical for modeling a PA with strong nonlinearities or long-term memory effects.

3. Simplified Volterra series based behavioral modeling

To reduce the complexity of the classical Volterra series, several simplified Volterra models have been proposed in recent years. In the following we discuss some of the most representative examples.

A. Preset Two-/Three-Box Models

In two-box models, we simply assume that the nonlinearity and memory effects of a power amplifier can be characterized separately, and then the PA can be represented by a single static nonlinearity with a linear dynamic block preceding or following it, which is called the *Wiener* model and the *Hammerstein* model, respectively [5].

If the linear dynamic block is represented by a Finite Impulse Response (FIR) filter, and the static nonlinearity block is represented by a power series, the Wiener Model can be written as

$$y_W(n) = \sum_{p=1}^{P} c_p \left(\sum_{i=0}^{M} h(i)x(n-i) \right)^p \quad (3)$$

and the Hammerstein Model can be represented by

$$y_H(n) = \sum_{i=0}^{M} h(i) \left(\sum_{p=1}^{P} c_p x^p(n-i) \right) \quad (4)$$

Compared to (1), we can see that the two-box models are just special cases of the general Volterra series. For example, in the Wiener model, the pth-order Volterra kernel can be formulated as

$$h_p(i_1, i_2, \cdots, i_p) = c_p h(i_1) h(i_2) \cdots h(i_p) \quad (5)$$

which indicates that the pth-order Volterra kernel of the Wiener model is equal to the product of p copies of the impulse response function of the linear block multiplied by the pth-order nonlinearity coefficient [5]. Similarly, the Volterra kernels of a Hammerstein system are only nonzero along their diagonals, and the diagonals of each of the kernels will be proportional to the impulse response of the linear subsystem, i.e.,

$$h_p(i_1, i_2, \cdots, i_p) = \begin{cases} c_p h(i) & i_1 = i_2 = \cdots = i_p \\ 0 & otherwise \end{cases} \quad (6)$$

A filter-nonlinearity-filter cascaded topology, i.e., a three-box model also is often used. By including a third box, we create an added dimension of flexibility which presumably can be used to improve on the two-box models. These filter and nonlinearity cascaded structures may be the simplest possible way to represent both memory and nonlinearity of a PA. However, in these models, the nonlinearity is normally obtained from a single-tone measurement at the center frequency. These models therefore can not predict interactions between the instantaneous tones, although it is well known that such interaction occurs in a real device. For example, the AM-AM and AM-PM characteristics may depend on the tone spacing in a wideband PA. Furthermore, the linear filter is normally identified by using small signal tests while memory effects of a PA may be present in the nonlinear region, such as those imposed by the bias networks.

To overcome these limitations, several improved models have been proposed. In [6], a polyspectral

technique is employed, in which the model consists of either a filter-nonlinearity or a nonlinearity-filter in parallel with a linear filter. Ku et al. [7] presented a parallel Wiener system whose parameters are extracted using two-tone measurements with different tone frequency spacings and power levels. In [8], an augmented Wiener model was proposed, in which a new parallel branch is added to the linear FIR filter to model memory effects more accurately. Although these enhanced models can be employed to characterize power amplifiers with reasonable accuracy, certain conditions have to be satisfied because they are based on specialized preset structures. Furthermore, since the model structures have been changed, the output of the model maybe no longer be linear with respect to the parameters which causes the model extraction to become more complicated in some cases.

B. Direct Pruning

In a practical situation, since nonlinear distortions and memory effects of an amplifier arise from different sources, not all of them have the same effects on the output. Some elements in the input vector may have less effect than others do, which leads to the result that some of the coefficients may be very small. It is reasonable to set them to zero during the model extraction and then remove them, which thereby simplifies the structure of the model and improves the simulation speed: this is called *Direct Pruning*.

Probably the simplest non-trivial pruned Volterra model is the diagonal Volterra model, also called the *memory polynomial model* [9]. In this model, all off-diagonal terms of the Volterra series are set to zero, i.e., $|i_m - i_n| = 0$ where i_m or i_n represents the delay of the input signal, then the Volterra model is simplified to

$$\tilde{y}(n) = \sum_{k=0}^{\frac{P-1}{2}} \sum_{i=0}^{M} a_{2k-1,i} \left| \tilde{x}(n-i) \right|^{2k} \tilde{x}(n-i) \qquad (7)$$

This model can be implemented by a block which is similar to a Finite Impulse Response (FIR) filter, but odd-order polynomials are used instead of the linear gain taps of the filter. In [10], the uniform time tap delays were replaced by non-uniform ones, i.e., sparse delays, to further reduce the number of coefficients. This diagonal restriction reduces the model complexity tremendously; however, it also has significant behavioral consequences, e.g. decreasing fidelity of the model, because, in some cases, the off-diagonal terms may be more important than the diagonal ones.

To improve the accuracy of the model, the all-zero off-diagonal restriction condition, $|i_m - i_n| = 0$, was relaxed in [11] to the "near-diagonality" restriction, i.e., $|i_m -$

$i_n| \leq l$, where l is a small integral. With the "near-diagonality" structural restriction, only the coefficients which are "far away" from the main diagonal in the model are removed. Although the number of coefficients is increased compared to the memory polynomial model, this solution gives us more flexibility on trading off the model accuracy and the model complexity, by selecting different l. For example, in [11], measurement results indicated that the model fidelity can be significantly improved if the second off-diagonals were included. Furthermore, this model can be implemented by a group of filter-banks, where the coefficients in the same diagonal line of the weight vector fall in the same FIR filter.

C. Modified/Dynamic Volterra series

The *Modified Volterra Series* [12-14], or *Dynamic Volterra Series* [15], is based on introducing the dynamic deviation function $e(n,i)$:

$$e(n,i) = x(n-i) - x(n) \qquad (8)$$

which represents the deviation of the delayed input signal $x(n-i)$ with respect to the current input $x(n)$. Substituting (8) in (1), the input/output relationship for a nonlinear system with memory can be described as

$$y(n) = y_s(n) + y_d(n) \qquad (9)$$

where $y_s(n)$ is the static part and can be expressed as a power series of the current input signal $x(n)$:

$$y_s(n) = \sum_{p=1}^{P} a_p x^p(n) \qquad (10)$$

in which a_p are the coefficients of the polynomial function, while $y_d(n)$ is the purely-dynamic part:

$$y_d(n) = \sum_{p=1}^{P} \sum_{r=1}^{p} x^{p-r}(n) \qquad (11)$$
$$\cdot \sum_{i_1=0}^{M} \cdots \sum_{i_r=0}^{M} w_{p,r}(i_1, \cdots, i_r) \prod_{j=1}^{r} e(n,i_j)$$

where $w_{p,r}(\cdot)$ represents the rth-order dynamic kernel of the pth order nonlinearity.

The most important property of this modified series is that it separates the purely static effects from the dynamic ones, which are intimately mixed together in the classical series. However, this modified Volterra series loses the property of linearity with respect to model parameters, which means that the output of the model is no longer linear with respect to the coefficients [14]. This leads to the consequence that models of this kind cannot be extracted in a direct way using established linear system estimation procedures

such as the least squares 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 [13-15].

In [16][17], the present authors extended the Modified Volterra Series to the discrete time domain, and rewrote it in the classical format after dynamic-order truncation. Then a new format of representation for the Volterra model was proposed as following,

$$
\begin{aligned}
y(n) &= \sum_{p=1}^{P} h_{p,0}(0,\ldots,0)x^p(n) \\
&+ \sum_{p=1}^{P}\{\sum_{r=1}^{p}[x^{p-r}(n)\sum_{i_1=1}^{M}\cdots\sum_{i_r=i_{r-1}}^{M} \\
&\cdot h_{p,r}(0,\cdots,0,i_1,\cdots,i_r)\prod_{j=1}^{r}x(n-i_j)]\}
\end{aligned}
\tag{12}
$$

in which the input elements are reorganized 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 was proposed, called *Dynamic Deviation Reduction* [16][17], in which higher order dynamics are removed since the effects of nonlinear dynamics tend to fade with increasing order in many real power amplifiers. 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 power amplifier 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.

D. Orthonormal/Orthogonal Projection

As mentioned earlier, the Volterra series has a clear nonlinearity structure which is a natural extension from a linear impulse response model, i.e., a transversal FIR filter. This FIR expansion may be not a very efficient description for a system with long-term memory. Because these finite impulse responses tend to decay linearly over time, the truncated "memory length" M directly depends on the duration of actual memory in the system. It is thus clear that M must be chosen large enough to include all "memories" which affect the output response of the system. Otherwise the approximation error would become too large and the dynamic representation of the model would be poor. This leads to the huge number of parameters that must be estimated in order to characterize a PA which has a very long memory. To address this issue, a recursive Infinite Impulse Filter (IIR) in [18] was used to construct the Volterra model more efficiently. However, due to its recursive structure, the system may become unstable in some cases.

In [19], an Orthonomal Basis Function (OBF), i.e., the Laguerre function, has been employed as the basis for the Volterra expansion. In the Laguerre-Volterra model, the Dirac impulses in the FIR filter, are replaced by the fixed-pole complex orthonormal Laguerre functions, which decay exponentially to zero at a controllable rate. The Laguerre functions have a similar structure as the IIR filer, but with a pre-decided fixed pole, so that the system can be stable. With the Laguerre functions based Volterra model, the long-term memory effects can be efficiently characterized [19]. Another orthonomal function, i.e., Kautz function, was also used in [20]. However, the choice of the orthonormal basis pole λ is critical. Usually, the pole is selected using *a priori* knowledge of the dominant dynamic of the system, such as, for example, considering the shape of its time or frequency response. An adequate choice can lead to a more efficient representation of the system since the better the pole choice, the faster will be the convergence of the series and the number of functions can be decreased.

On the other hand, the nonlinearity part of Volterra series is expanded from a polynomial function, in which the contributions of each of the system's Volterra kernels can not be separated. This leads to a situation where the coefficients estimated from different samples of the same signal may be different because the input vector contains various powers of the same signal samples. Wiener overcame this problem in his classical theory of nonlinear systems [2] by using the so-called G-functions, which are orthogonal when the input is a white Gaussian process. Using orthogonal signals for representing nonlinear systems allows the coefficients of each signal to be evaluated independently of every other signal and then the individual Volterra kernels can be isolated. Unfortunately, a white Gaussian signal is difficult to construct in real microwave measurements.

Alternatively, a multi-sine of randomized phases can be employed to build an orthogonal model in the frequency domain [21]. However, this model still inherits the high complexity of general Volterra models because no effort has been made to simplify the underlying model structure.

4. Conclusions

A short overview of various simplified Volterra series based behavioral models for RF/microwave power amplifiers has been presented in this paper. Both advantages and limitations of different models have been discussed. In practical applications, model structure and pruning algorithm selection depends on the characteristics of the real system, model fidelity requirements and various specific conditions. Both the complexity of model structure and the feasibility of model extraction have to be considered in deciding which model to use.

References

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

[2] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, R. E. Krieger Publishing, reprint ed., 1989.

[3] V. J. Mathews, G. L. Sicuranza, *Polynomial Signal Processing*, John Wiley & Sons, 2000.

[4] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems*, Kluwer Academic Publishers, 2nd ed., 2000.

[5] D. T. Westwick, R. E. Kearney, *Identification of Nonlinear Physiological Systems*, IEEE Press, 2003.

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

[7] H. Ku, M. Mckinley, and J. S. Kenney, "Quantifying memory effects in RF power amplifiers," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, no.12, pp. 2843-2849, Dec. 2002.

[8] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "De-embedding static nonlinearities and accurately identifying and modeling memory effects in wide-band RF transmitters,", *IEEE Trans. on Microwave Theory and Techniques*, Vol. 53, no. 11, pp. 3578 – 3587, Nov. 2005.

[9] J. Kim, K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, no. 23, pp. 1417-1418, Nov. 2001.

[10] H. Ku, J. S. Kenney, "Behavioral modeling of nonlinear RF power amplifiers considering memory effects", *IEEE Trans. on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2495-2504, Dec. 2003.

[11] A. Zhu, T. J. Brazil, "Behavioral Modeling of RF Power Amplifiers Based on Pruned Volterra Series", *IEEE Microwave and Wireless Components Letters*, Vol. 14, pp. 563-565, Dec. 2004.

[12] F. Filicori, G. Vannini, "Mathematical approach to large-signal modeling of electron devices," *Electronics Letters*, vol. 27, no. 4, pp. 357-359, 1991.

[13] D. Mirri, G. Luculano, et al, "A modified Volterra series approach for nonlinear dynamic systems modeling," *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 8, pp.1118-1128, Aug. 2002.

[14] D. Mirri, F. Filicori, G. Iuculano, G. Pasini, "A nonlinear dynamic model for performance analysis of large-signal amplifiers in communication systems," *IEEE Trans. on Instrumentation and Measurement*, vol. 53, pp. 341 – 350, Apr. 2004.

[15] E. Ngoya, N. L. Gallou, et al, "Accurate RF and microwave system level modeling of wideband nonlinear circuits," *IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 1, pp. 79-82, 2000.

[16] A. Zhu, J. Dooley, and T. J. Brazil, "Simplified Volterra Series Based Behavioral Modeling of RF Power Amplifiers Using Deviation-Reduction", *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 1113-1116, 2006.

[17] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic Deviation Reduction Based Volterra Behavioral Modeling of RF Power Amplifiers", *IEEE Trans. on Microwave Theory and Techniques*, to appear, Dec. 2006.

[18] J. Dooley, B. O'Brien, and T. J. Brazil, "Behavioral Modeling of RF Power Amplifiers Using Adaptive Recursive Polynomial Functions", *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 852-855, Jun. 2006.

[19] A. Zhu, T. J. Brazil, "RF Power Amplifiers Behavioral Modeling Using Volterra Expansion with Laguerre Functions," *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 963-966, 2005.

[20] M. Isaksson, D. Rönnow, "A Kautz-Volterra Behavioral Model for RF Power Amplifiers," *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 485-488, 2006.

[21] P. M. Lavrador, J.C. Pedro, and N.B. Carvalho, "A new Volterra series based orthogonal behavioral model for power amplifiers," *The Asia-Pacific Microwave Conference Proceedings*, Dec. 2005.