General Nonlinear Feed-forward RF Model for Power Amplifiers

Telmo R. Cunha(1), José C. Pedro(1), Pedro M. Cabral(1), and Anding Zhu(2)

(1)Instituto de Telecomunicações – Universidade de Aveiro, Aveiro, 3810-193, Portugal
(2)University College Dublin, Dublin, 4, Ireland

***Abstract* — This paper presents a procedure to design a non-recursive model, suited for simulation and able to model short- and long-term memory effects, that is the best nonlinear approximator, up to a certain polynomial order, of a feedback behavioral model whose topology was derived from the PA physical behavior. A simple and systematic model extraction procedure, based on single- and two-tone tests, is given for a model truncated to 3rd order, and its validation is demonstrated through ADS simulation comparison.**

***Index Terms* — Feedforward systems, modeling, nonlinear systems, power amplifiers, Volterra series.**

I. INTRODUCTION

Most of power amplifiers, PAs, behavioral models have been derived for representing only the input-output impairments suffered by the slowly varying information signal, i.e., the complex envelope of the modulated RF excitation. These are the so-called low-pass equivalent models. Unfortunately, as they were especially conceived for wireless system simulation at the complex envelope level, the models suffer from two important limitations.

First, they can neither account for any change on the input or output impedance mismatches, nor any change on the out-of-band PA terminations, mostly at the high-frequency second harmonics and the low-frequency envelope components. Their second drawback is that such complex envelope models are system level representations that can not be incorporated in a nonlinear circuit simulator. In fact, the handled signals are the input and output low-pass complex envelopes, not the true band-pass real RF modulated signals.

An accurate circuit level behavioral model should be capable of simultaneously representing the short-term memory effects of the PA's active device and input/output matching networks, and the long-term memory effects usually attributed to the bias circuitry, the transistor's dynamic electro-thermal properties and (possibly) semiconductor charge-carrier trapping. Nevertheless, despite no one denies the practical relevance of this behavioral modeling problem, it still constitutes an open field of research.

Curiously, there is no theoretical impediment to build such a model. The problem is, in fact, of practical nature. Trying to rely on systematic procedures for model extraction and on the guaranteed stability provided by feed-forward (or non-recursive) models, microwave engineers have preferred these structures in detriment of their feedback (or recursive) counterparts.

II. MODELING COMPLEXITY VERSUS GENERALITY

Using nonlinear models of guaranteed generality, such as Volterra series or time-delay neural networks [1]-[2], to simultaneously represent short- and long-term memory effects has, in terms of simulation, severe drawbacks. If the product between sampling frequency times the duration of the long-term effects is high, the time-domain general representations must contain a huge number of parameters which makes them impracticable for simulation purposes, not to mention the required extraction effort. For instance, the 3rd order kernel of a general Volterra series modeling a system whose fundamental frequency is f_c would require a time resolution lower than $1/(6f_c)$. For illustrative purposes, say that f_c=900MHz. If a 3rd-order long-term effect of 1µs was to be modeled, the 3rd order time-domain kernel would contain $157{,}464\cdot 10^9$ parameters, and all of them would be necessary to determine the system response to a certain excitation.

Schemes, where different time resolutions are considered for short- and long-term effects, can only be implemented if both effects are clearly separated by the model. Otherwise, the representation could not be that of an (multi-dimensional) impulse response of a physical system. For the model to impose that effect separation, its topology has to be different from its general form. That is, shaping the model topology makes it less general, but can drastically decrease its implementation complexity.

Fortunately, transistors and PAs do not seem to be general nonlinear systems [3], so modeling them with general structures implies that part of the model parameters (which, like all others, are necessary for simulation) generate a combined output contribution which is null. This encourages the use of a model topology in detriment of general models. In fact, the topology of a model contains, by itself, a considerable amount of information which, in the case of a general structure, would be hidden in its parameters' values (and their combinations). The main difficulty is in the tailoring of the model topology, which cannot be too restrictive (it must model general transistors and PAs) nor too general (to reduce implementation complexity).

If this model tailoring is based on the physical knowledge of the transistor behavior, its predictive capability is automatically guaranteed.

1-4244-0688-9/07/$20.00 ©2007 IEEE

III. Transistor Feed-forward Model Design

In [3], it is proved that the PA's active device can be modeled by a recursive model where a memoryless polynomial is subjected to a linear memory feedback (Fig. 1).

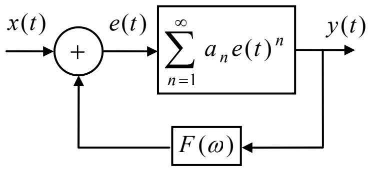

Fig.1. Transistor model derived from physical behavior [3].

As is shown from nonlinear system theory, the best approximation of a nonlinear system up to a certain order is that resulting from the truncation of its Volterra series up to the kernel of that same order. So, to build the feed-forward model topology that best approximates that of Fig. 1, the Volterra kernel expressions were obtained for the latter (up to the 5th order), and for each kernel, the respective feed-forward topology that produces the exact same kernel expression was determined.

For simplicity reasons, the following analysis considers only a model truncation to 3rd order. However, it can be directly extended to higher orders.

Fig. 2 shows the resulting topology, up to 3rd order.

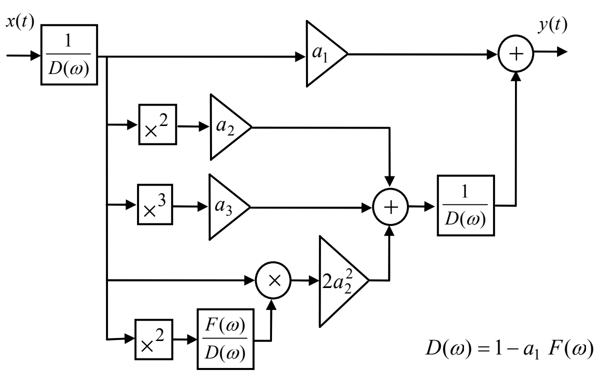

Fig. 2. Feed-forward PA model up to the 3rd order kernel.

The complete feed-forward transistor model, up to order n, would then be the parallel of the models of all the kernels up to order n. It is interesting to notice that the structure of the kernel models is basically the same. All (except for the 1st order one, for evident reasons) have the same input and output linear FIR filter $1/D(\omega)$ and, in between, is a varying number of parallel branches producing all possible combinations of the input that would generate terms of order n. The filter appearing in the parallel branches (in those that have memory) is always $F(\omega)/D(\omega)$.

If no restrictions are imposed to this model, a single tone test, swept from DC to a frequency as high as desired, would be enough to extract the whole model since all filters are derived from only one filter – $F(\omega)$ (which can be determined from a small-signal analysis, from $1/D(\omega)$).

Let us consider, now, that an input matching network imposes that $x(t)$ is restricted to the fundamental band, but the output is not restricted. In this case, a single tone test (swept in the fundamental band, but not restricted to small-signal) would still suffice for the extraction of the model filters. For example, $F(2\omega)$ (where ω is at the fundamental) can be extracted from the output at the second harmonic, giving thus the dynamic information required for the 2nd order branch. Then, this $F(2\omega)$ is also used in the third order branch as the filter following the squarer. A similar reasoning can be applicable for the base-band of $F(\omega)$.

If the system is also band-limited at the output, by means of an output matching network, it would not be possible to extract any information outside the fundamental band. Therefore, the model extraction would require an input signal more elaborated than a single-tone. In fact, the problem imposed by the band-pass filters is one that, in control theory, is known as system observability. Note that, as it is now impossible to directly observe some characteristics of the transistor behavior (e.g. second order behavior that only produces out-of-band behavior), they can be deleted from the model. The following sections present a model and a systematic extraction procedure for this more difficult band-limited case.

IV. Model Reduction

Looking into the time-domain Volterra kernels (the ones used for simulation) of the topology of Fig. 2, it is seen that these will be fully populated with values that can be non-zero (within the memory span). This is due to the existence of a cascaded sequence: *linear filter → nonlinearity → linear filter*. In this case, there would be no reduction of the number of degrees-of-freedom that could be used to significantly simplify the extraction procedure when compared to the general case.

Fortunately, these devices are usually designed so that their linear behavior is approximately flat at the fundamental band. Considering such flatness assumption in the model of Fig. 2, i.e. assuming $F(\omega)$ and $D(\omega)$ are flat, further simplifications are obtained. Fig. 3 presents this simplified version of the topology of Fig. 2. Not represented is a constant time delay associated to the device transfer function. Also, reflected in Fig. 3, is the elimination of redundant parameters of the model of Fig. 2. For example, the factor $2a_2^2$ was incorporated into the filter $F(\omega)/D(\omega)$ following the squarer, so was the cubic operator branch (since it has no memory associated), resulting the new filter $G(\omega)$.

Also worth to mention in Fig. 3 is the separation of filter $G(\omega)$ into the parallel of filters $G_{bb}(\omega)$ and $G_{2nd}(\omega)$. As the signal entering the squarer is always at the fundamental band, the squarer output will only have components at base-band and at the 2nd harmonic band.

At this stage it is interesting to notice the topological similarities between the model of Fig. 3 and those presented in [4]-[5], which shows how this work can be used to give a physical justification for such apparently arbitrary topologies, and also, as we will discuss next, provide them with a systematic extraction procedure.

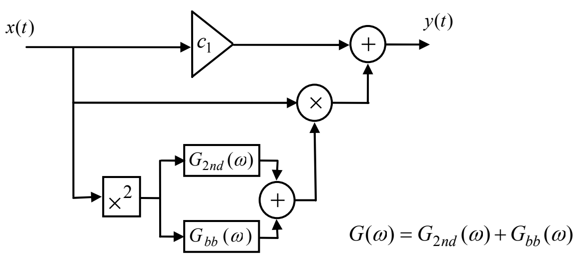

Fig. 3. Up to 3rd order model with flatness assumption at the fundamental band, and with minimal representation.

The kernel expressions for the topology of Fig. 3, in frequency- and time-domain are (1) and (2), respectively.

$$\begin{cases} S_1(\omega) = c_1 \\ S_3(\omega_1,\omega_2,\omega_3) = \frac{1}{3}\left(G(\omega_1+\omega_2)+G(\omega_1+\omega_3)+G(\omega_2+\omega_3)\right) \end{cases} \quad (1)$$

$$\begin{cases} h_1(\tau) = \begin{cases} c_1 & ,\tau = 0 \\ 0 & ,otherwise \end{cases} \\ h_3(\tau_1,\tau_2,\tau_3) = \begin{cases} g(0) & ,\tau_1=\tau_2=\tau_3=0 \\ \frac{1}{3}g(\tau) & ,\begin{vmatrix} \tau_1=\tau_2=\tau\neq 0 \wedge \tau_3 = 0 \\ \tau_1=\tau_3=\tau\neq 0 \wedge \tau_2 = 0 \\ \tau_2=\tau_3=\tau\neq 0 \wedge \tau_1 = 0 \end{vmatrix} \\ 0 & ,otherwise \end{cases} \end{cases} \quad (2)$$

It is clear from (2) that the 3rd order time-domain Volterra kernel is now reduced to the main diagonals of the kernel supporting planes. All other kernel parameters are null. This constitutes a significant model simplification, both for extraction and simulation.

V. Model Extraction

Considering the model of Fig. 3, with the assumption that the PA behaves as a flat linear filter at the fundamental band, under linear operation, it will be shown that its parameters can be uniquely determined through a sequence of single-tone tests followed by two-tone ones.

Under a small-signal analysis, only the upper branch of the model will be relevant for the output. So, c_1 is obtained from a single-tone test, which is also used to determine the system time-delay.

Then, for the input power level at which the PA will be normally working, the following equal amplitude two-tone test is performed. One of the tones is kept with the same frequency (say, ω_1) for all iterations, while the second tone (with frequency ω_2) is swept through the fundamental band. The IMD output components at frequencies $2\omega_1$-ω_2 and $2\omega_2$-ω_1 (the adjacent channel components) are determined through (3)-(6), where real and imaginary parts were split.

$$G_R(2\omega_1) + 2\,G_R(\omega_1-\omega_2) = \frac{Y_R(2\omega_1-\omega_2)}{X(\omega_1)^2\,X(\omega_2)^*} \quad (3)$$

$$G_I(2\omega_1) + 2\,G_I(\omega_1-\omega_2) = \frac{Y_I(2\omega_1-\omega_2)}{X(\omega_1)^2\,X(\omega_2)^*} \quad (4)$$

$$G_R(2\omega_2) + 2\,G_R(\omega_1-\omega_2) = \frac{Y_R(2\omega_2-\omega_1)}{X(\omega_2)^2\,X(\omega_1)^*} \quad (5)$$

$$G_I(2\omega_2) - 2\,G_I(\omega_1-\omega_2) = \frac{Y_I(2\omega_2-\omega_1)}{X(\omega_2)^2\,X(\omega_1)^*} \quad (6)$$

Considering the sum of equations (4) and (6) and with the two-tones set at very close frequencies, the imaginary part of the 2nd harmonic band of filter $G(\omega)$ can be determined with considerable precision from (7).

$$\left| \begin{array}{l} G_I(2\omega_1) \approx G_I(2\omega_2) \approx \frac{1}{2}\left(\frac{Y_I(2\omega_1-\omega_2)}{X(\omega_1)^2\,X(\omega_2)^*} + \frac{Y_I(2\omega_2-\omega_1)}{X(\omega_2)^2\,X(\omega_1)^*}\right) \\ as\ \omega_2 \to \omega_1 \end{array} \right. \quad (7)$$

Now, keeping ω_1 constant, and sweeping ω_2, the imaginary part of the base-band region of $G(\omega)$, $G_I(\omega_1$-$\omega_2)$, can be extracted from (4), since $G_I(2\omega_1)$ is known.

A similar procedure for the real part of $G(\omega)$ would not be successful since physical constraints impose $G_R(\omega)$=$G_R(-\omega)$.

So, three unknowns are left for only two equations. From (3) and (5) it is not possible to dissociate $G_R(2\omega_1)$ [or $G_R(2\omega_2)$] from $2G_R(\omega_1$-$\omega_2)$. Actually, this is due to the fact that, as in Fig. 3, base-band and 2nd harmonic bands of $G(\omega)$ are topologically in parallel. This means that there is an extra degree-of-freedom associated to the real parts of both bands, but the restriction of physical realization of the filters imposes, for example, $G_R(0)$.

With $G_R(0)$ fixed, and using the two-tone test where ω_1 and ω_2 are very close, a good estimate of $G_R(2\omega_1)$ can be obtained from (3). Then, the sweeping of ω_2 allows the determination of $G_R(\omega_1$-$\omega_2)$, also from (3). Finally, $G_R(2\omega_2)$, the remaining unknown, is obtained from (5).

VI. Model Validation

A complete PA, with input and output matching networks, and with short- and long-term memory behavior, was simulated in the ADS software from Agilent Technologies, Inc., having a central frequency of 900MHz. The CW and two-tone extraction procedure described in section V, at four distinct input power levels, was executed in that software, and its 3rd order models were obtained (one for each power level).

The extracted model was then validated against simulations of the response to a CDMA signal, with the same four power levels.

In Table I are the obtained NMSE figures of the comparison between the ADS and feed-forward model generated complex envelopes of the output. As expected, the NMSE for the first two power points is very high because the system is mostly linear. As nonlinear components get more relevant, the NMSE decreases, indicating that higher than 3rd order behavior becomes non-negligible at the output. In Fig. 4, a plot of part of the time-domain complex envelope (real and imaginary parts) is depicted. Fig. 5 shows the frequency-domain complex envelope comparison.

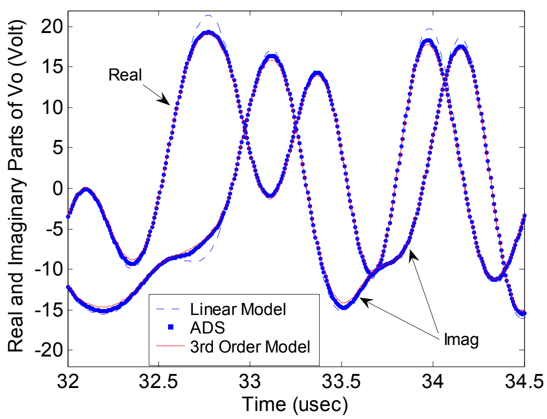

Fig. 4. Time-domain complex envelope comparison (P_{in}=15dBm).

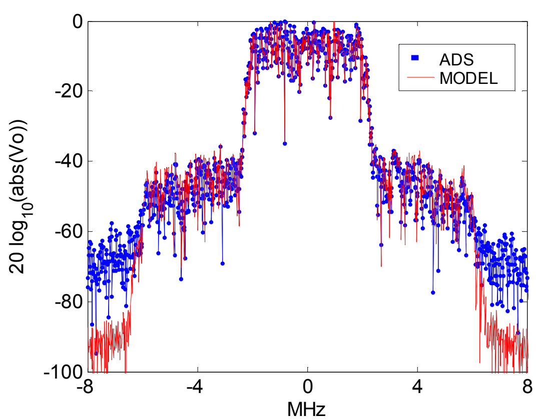

Fig. 5. Output voltage complex envelope spectrum comparison (P_{in}=15dBm).

These results show that the extracted model provide accurate prediction of the linear and nonlinear behavior up to 3rd order. To consider higher order contributions, more parallel branches have to be added to the topology of Fig. 3, and a more complex extraction procedure must be performed.

TABLE I
NMSE FIGURES FOR DIFFERENT INPUT POWER LEVELS

P_{in} (dBm)	10	12	15	18
Compression Point (dB)	0.09	0.12	0.30	0.94
NMSE (dB)	-40.6	-41.2	-29.5	-24.5

VII. CONCLUSIONS

This paper indicates an analytical procedure to construct the non-recursive RF model of the PA's active device, which best approximates, up to a desired polynomial order, its feedback behavioral model deduced in [3]. In the above text, an exact extraction procedure based on single- and two-tone inputs is presented for a model truncation of 3rd order. With simulated ADS data, the obtained model is validated, confirming its good predictive properties.

Extension of the presented analysis to higher orders is straight-forward although laborious. The 3rd order model is sufficient to illustrate the presented feed-forward model design, without adding extra topology complexity of higher order models.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support provided by the Institute of Telecommunications – IT and F.C.T. (Portuguese National Science Foundation) through the Project ModEx and the EC under the Network of Excellence TARGET.

REFERENCES

[1] J. C. Pedro, and S. A. Maas, "A comparative overview of microwave and wireless power amplifier behavioral modelling approaches", *IEEE Trans. on Microwave Theory and Tech.*, vol. MTT-53, pp.1150-1163, Apr. 2005.

[2] S. Boyd and L. Chua, "Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series", *IEEE Trans. on Circuits and Syst.*, vol. CAS-32, pp.1150-1161, Nov. 1985.

[3] J. C. Pedro, N. B. Carvalho, and P. M. Lavrador, "Modeling nonlinear behaviour of band-pass memoryless and dynamic systems", *IEEE MTT-S Int. Microwave Symp. Dig.*, Philadelphia, PA, pp. 2133-2136, Jun. 2003.

[4] C. P. Silva, A. A. Moulthrop, M. S. Muha, and C. J. Clark, "Application of polyspectral techniques to nonlinear modeling and compensation," *IEEE MTTS Intl. Microwave Symposium Dig.*, vol. 1, pp. 13–16, Phoenix, Arizona, May 2001.

[5] A. Walker, M. Steer, K. Gard, and K. Gharaibeh, "Multi-slice behavioral model of RF systems and devices", *Proc. IEEE Radio and Wireless Conf.*, pp.71-74, Sep. 2004.