FOCUSED ISSUE FEATURE

Behavioral Modeling and Predistortion

Fadhel M. Ghannouchi and Oualid Hammi

© COMSTOCK

Wireless communication systems are continuously evolving to support more users and provide higher data rates within the limited and overcrowded radio-frequency (RF) spectrum. To ensure spectral efficiency, compact constellations and multiple access techniques based on code division multiple access (CDMA) and orthogonal frequency division multiplexing (OFDM) are used. Yet, this spectral efficiency is achieved at the expense of significantly tougher requirements on the performance of the RF front end. Indeed, the resulting nonconstant-envelope, amplitude-modulated signals unavoidably stimulate the transmitters' nonlinearities. This nonlinear behavior is mainly generated by the power amplifier (PA) operating close to saturation for power efficiency considerations. Operating the PA in linear mode with such signals, which have high peak-to-average power ratio (PAPR), will result in significantly low power efficiencies. Accordingly, the design of the power amplification stage is often perceived as being driven by a tradeoff between linearity and power efficiency.

Current state-of-the-art power amplification systems often use a Doherty PA for

Fadhel M. Ghannouchi (fadhel.ghannouchi@ucalgary.ca) is a Distinguished Microwave Lecturer of the MTT Society and a Fellow of the IEEE. Oualid Hammi (ohammi@gradiotech.com) is with Green Radio Technologies Inc., Calgary, AB, Canada and is a Member of the IEEE. They are both with iRadio Laboratory, Electrical Engineering Department, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada.

Digital Object Identifier 10.1109/MMM.2009.934516

high efficiency at output power backoff, and a digital predistorter to restore the required linearity performance. Digital predistortion is currently the preferred linearization technique and is widely used for applications with, typically, up to 20 MHz bandwidth. Digital predistortion consists of applying a complementary nonlinearity upstream of the PA so that the cascade of the digital predistorter and the PA behaves as a linear amplification system. In this context, behavioral modeling is crucial to predict the nonlinearity of the PA in particular and the transmitter in general. Moreover, behavioral modeling is even more important since predistortion can also be considered as a behavioral modeling problem. The synthesis of the predistortion function is equivalent to the behavioral modeling of the PA's reverse function obtained by swapping the PA's input and output signals with appropriate small-signal gain normalization.

The key advantage of behavioral modeling resides in the fact that it does not require deep knowledge of the RF circuit physics and functionality. As illustrated in Figure 1, behavioral modeling simplifies the modeling of the RF circuit to the identification of a mathematical formulation that relates the input and the output of the device under test (DUT) that can be considered as a black box. Accordingly, behavioral modeling appears as a time- and resource-efficient process for transmitter performance evaluation and digital predistorter design. The performance of behavioral modeling is influenced by two key aspects, the observation and the formulation.

- The observation refers to the accurate acquisition of the signals at the input and output of the DUT while exciting the appropriate behavior.
- The formulation corresponds to the choice of a suitable mathematical relation that describes all the significant interaction between the DUT's input and output signals.

Accordingly, a limited a priori knowledge of the DUT is required to observe the right behavior and choose the adequate formulation.

This article complements previous work overviewing and comparing behavioral models [1], [2]. It addresses the major issues related to the state of the art of behavioral modeling and digital predistortion of PAs for 3G+ wireless communication infrastructures. In this article, digital predistortion is presented from a behavioral modeling perspective with an emphasis on the similarity between the behavioral modeling and the digital predistortion processes. Thus, this article focuses on distortion cancellation based on digital predistortion techniques and does not cover other linearization techniques. First, various key aspects that need to be considered for accurate observation of the DUT behavior are discussed. Then, a thorough review of state-of-the-art behavioral models is presented. Model performance assessment

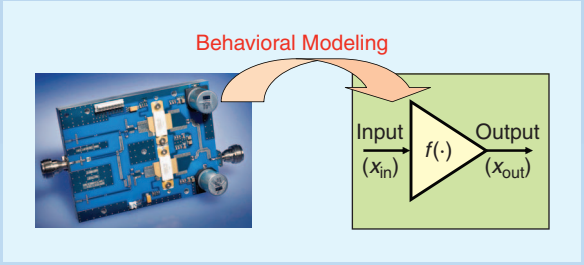

Figure 1. *Black-box based behavioral modeling reduces the modeling of a complex analog circuit into a mathematical formulation that relates its input to its output.*

is discussed and a comparison between state-of-the-art models performance and complexity is carried out. Finally, a software solution for closed-loop digital predistortion is briefly introduced.

Digital Predistortion from a Behavioral Modeling Perspective

The block diagram of a digital-predistortion-based linear amplification system is presented in Figure 2. The signal at the input of the digital predistorter ($x_{in\_DPD}(n)$) can be derived from that at the output of the PA stage according to

$$x_{in\_DPD(n)} = \frac{x_{out}(n)}{G}, \tag{1}$$

where $x_{out}(n)$ and G are the output waveform and the gain of the linearized PA, respectively.

Accordingly, while the behavioral modeling of the DUT consists of identifying the function f_{DUT} that satisfies $f_{DUT}(x_{in}(n)) = x_{out}(n)$, the synthesis of the digital predistortion function is equivalent to the estimation of the function f_{DPD} such that $f_{DPD}(x_{out}(n)/G) = x_{in}(n)$. Two approaches can be considered to synthesize the predistortion function. The first consists of synthesizing and then inverting the model of the DUT to be linearized. The second approach consists of identifying the digital predistortion function directly since the input and output waveforms of the digital predistorter can be calculated from the measured input and output waveforms of the DUT, as illustrated in Figure 2. This approach is often referred to as the indirect learning technique initially proposed in [3] and is widely used in digital predistortion systems.

Several digital predistortion functions can be selected, depending on the choice of the digital predistorter's small-signal gain, which sets the value of

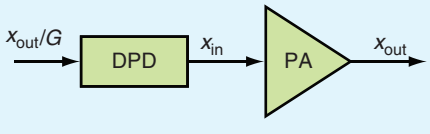

Figure 2. *Block diagram of digital predistortion (DPD) based linear power amplifier (PA).*

When memory effects, also known as dynamic distortion, are present, the instantaneous complex gain of the DUT is a function of the actual input sample and a finite number of the preceding input samples.

the linearized PA gain (G) [4]–[6]. However, the choice of the digital predistorter's small-signal gain does not affect the PA's efficiency. A comprehensive study on the selection of the digital predistorter's small-signal gain was recently undertaken in [6].

Observation

What Needs to Be Observed?

Base station PAs are designed to handle multicarrier nonconstant-envelope, amplitude-modulated signals with bandwidths spanning 20–60 MHz. Under such conditions, the nonlinear behavior of PAs/transmitters has two major components—the static nonlinearity and the memory effects. The static nonlinearity, also referred to as the memoryless nonlinearity, corresponds to the distortion generated by the DUT when memory effects are not present. In such case, the instantaneous complex gain of the DUT is a function of the actual input sample only (assuming that there is no propagation delay through the DUT) and not the preceding input samples. Memory effects, also known as dynamic distortions, can be either thermal or electrical. When they are present, the instantaneous complex gain of the DUT is a function of the actual input sample (again assuming that there is no propagation delay through the DUT) and a finite number of the preceding input samples. The contribution of the static nonlinearity to the overall behavior of the DUT is notably stronger than that of the memory effects. However, both the static nonlinearity and the memory effects are equally important, especially in digital predistortion, because, in most cases, they both need to be compensated for to meet the linearity requirements for 3G+ systems. Thus, a major challenge is to accurately identify both behaviors while taking into account their interdependence.

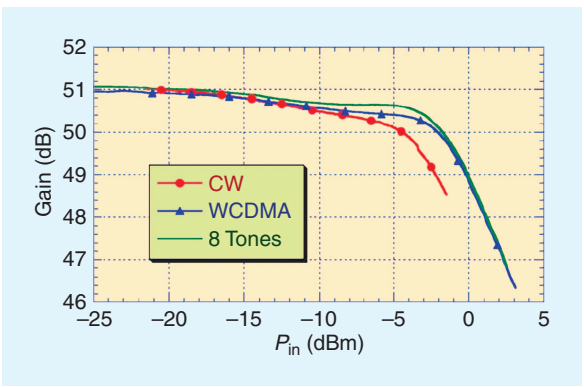

Figure 3. *Effects of test stimuli on measured AM/AM characteristics: The measured AM/AM characteristic varies with the type of stimuli. (From [14].)*

The memory effects and the static nonlinearity of PAs and transmitters are related and depend on the operating conditions of the DUT. These conditions are device dependant (bias condition, temperature, etc…) and signal dependant (signal statistics, average power, and bandwidth). In this section, the focus is on the DUT behavior dependency with respect to the signal characteristics. The bias conditions and operating temperature also need to be controlled for accurate and repeatable observation of the DUT behavior.

Several stimulus signals have been used over the years to characterize the behavior of PAs. These stimuli include continuous wave (CW), two-tones, multitone, and the actual test signal (CDMA, OFDM, etc…) [7]–[14]. The limitations of CW-based characterization are mainly attributed to the narrowband nature of this stimulus and its constant envelope power.

Two-tone and multitone signals have controllable bandwidths that can be set by the spacing between the tones. They offer a good basis for mathematical analysis through the approximation of the CDMA and OFDM spectra by a discrete spectrum for which each frequency component can be analytically derived [15]–[17]. The PAPR of multitone signals can be controlled by a proper distribution of the tones' phases for an accurate approximation of the real signal's statistics. The main problem in two-tone-based measurements is that the power distribution over the signal bandwidth is uneven. The energy of two-tone signals is concentrated within very narrow bandwidths at the edges of the frequency band spanning between these two tones. Conversely, when more realistic signals (such as single or multicarrier CDMA or OFDM type of signal) covering the same frequency bandwidth are used, the power distribution over the frequency band is not the same as that of two-tone signal. For such realistic signals, the power is evenly distributed over more bandwidth within the same frequency band. Such a difference in the power distribution is likely to lead to different behaviors of the DUT. For example, memory effects caused by a four-carrier wideband CDMA (WCDMA) signal with 20 MHz bandwidth is stronger when the carrier configuration is 1001 compared to the case where the carrier configuration is 1111. As an extension, it can be predicted that a two-tone signal would lead to an overestimation of the memory effects of the DUT. Thus, the most appropriate approach consists of applying the real signal to characterize the DUT behavior.

Figure 3 presents the static nonlinearity of a Class AB PA measured under various drive signals. Three

types of signals were used to measure the AM/AM characteristic of this DUT. These signals include a CW signal, an eight-tone signal, and a WCDMA signal. The eight-tone and WCDMA signals excite comparable behaviors of the DUT. The slight difference between the two observed behaviors is mainly due to the variation in the average operating power during the measurements. Conversely, the CW-based measurement led to a more pronounced gain compression compared to that measured using the two previous signals. This is mainly attributed to the extra self-heating that occurs at high input power levels with CW measurements. This self heating is due to the limited sweep time speed. This figure illustrates the dependency of the observed behavior on the stimuli [14].

Modern communication signals have three parameters that might influence the behavior of PAs. These are the signal's average power, the signal's statistics (PAPR and probability density function), and the signal's bandwidth. Typically, the behavior of a PA is sensitive to variations in the operating average power (even less than 1dB in some cases) [18]. Conversely, it is almost insensitive to the signal statistics (typically up to 3–4 dB variation in the signal's PAPR does not affect the behavior of a PA as long as the signals' statistics follow the same distribution) [18]. These signal properties mainly influence the static nonlinearity of the DUT. On the other hand, the signal bandwidth has a direct impact on the memory effects exhibited by the DUT and have a second-order effect on the static nonlinearity [19], [20]. An example of the dependence of the PA behavior to the characteristics of the test signal is reported in Figure 4 for a Doherty amplifier prototype. In Figure 4(a), the AM/AM characteristics of the DUT were measured at constant average power and bandwidth for WCDMA signals having different PAPR. These signals were generated by applying a crest factor reduction algorithm with variable clipping ratio to the same WCDMA input signal. The operating average power was chosen so that the DUT is driven up to (but not beyond) its input saturation power when the signal having the highest PAPR is used. This illustrates the quasi-insensitivity of the DUT behavior to such changes in the signal statistics. The same signals were used to measure the behavior of the DUT operated at various average power levels. In this second test, the operating average power of each signal was set so that the PA operated at a constant peak power in all cases. The measurement results show that the AM/AM characteristics of the DUT vary with the operating average power as illustrated in Figure 4(b). These results clearly show the sensitivity of the DUT to the

The effect of nonlinearities in wireless transmitters is a critical issue when nonconstant-envelope modulated signals are being used.

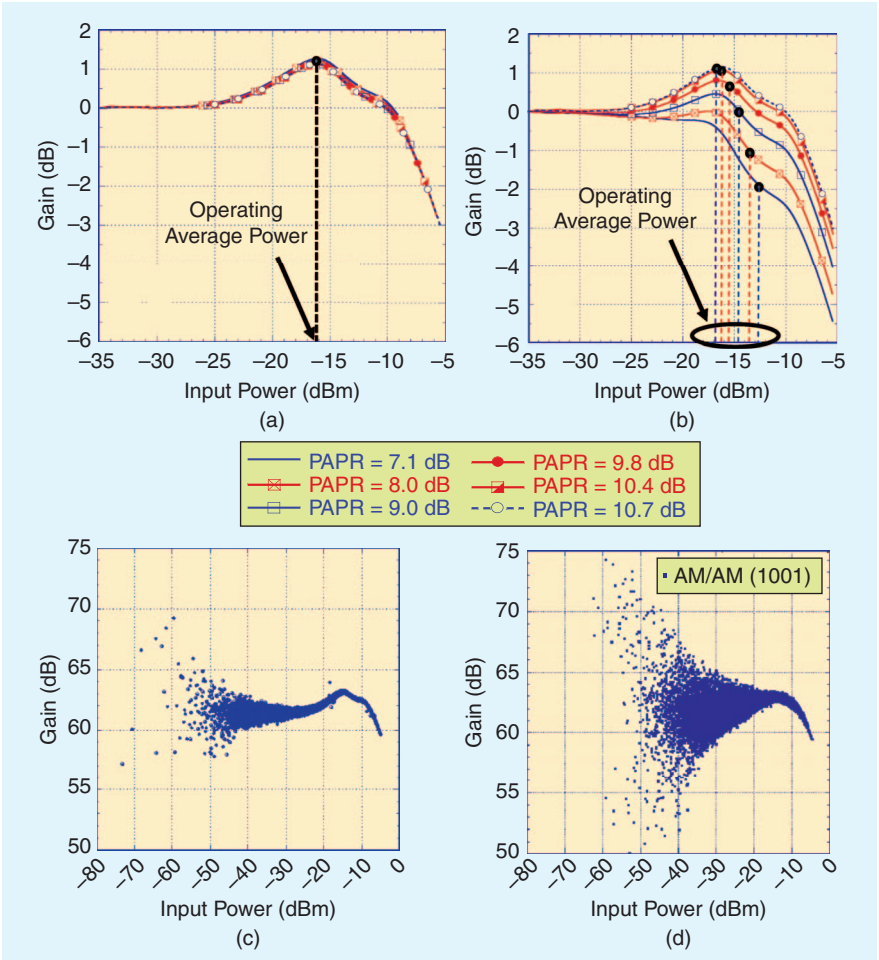

Figure 4. *Measured AM/AM characteristics of a Doherty prototype versus WCDMA signal statistics. (a) The behavior of the device under test is almost insensitive to variations in the peak-to-average-power ratio of the excitation. (b) The behavior of the device under test varies with the operating average power. (c) Measured sensitivity to bandwidth (AM/AM under a one-carrier WCDMA input signal). (d) The measured sensitivity to bandwidth [AM/AM under a four-carrier WCDMA input signal shows more dispersion in comparison to (c)]. [(a) and (b) from [18], (c) and (d) from [20].]*

Identifying and modeling nonlinear behavior is an important step in designing the circuitry and algorithms that will adequately compensate for it.

signal's average power. Finally, the memory effects of the DUT were qualitatively evaluated at a constant average power for signals having similar statistics but different bandwidths [Figure 4(c) and (d)]. For this last test, the PA was successively driven by a single carrier WCDMA signal and then a four-carrier WCDMA signal having a 1001 carrier configuration. Quantitative evaluation of memory effects intensity can be performed using several metrics such as those presented in [20]–[23].

How Should It Be Observed?

The behavior of the DUT can be experimentally measured using the instantaneous complex baseband waveform approach. A typical setup is presented in Figure 5. The digital baseband waveform is downloaded from an arbitrary waveform generator that will feed the PA with the corresponding RF modulated signal. At the output of the amplification stage, the signal is processed with a vector signal analyzer (VSA) that performs the signal down conversion and digitization. The digital signal, considered as the DUT output signal, is then acquired from the VSA for processing. In such experimental systems, the DUT input signal can be considered either at the input or at the output of the arbitrary waveform generator. In the first case, the DUT corresponds to the entire transmitter, including the arbitrary waveform generator as well as the amplification stage. While in the latter case, the arbitrary waveform generator is not part of the system being modeled.

To accurately observe the behavior of the DUT, the experimental setup needs to satisfy several constraints, mainly related to the dynamic range and the bandwidth. The experimental setup in both the signal generation and the signal acquisition paths should have a dynamic range that provides an adjacent channel power ratio (ACPR) that is better than the targeted linearity performance. The system's ACPR sets the upper limit for the digital predistorter's performance. Moreover, the bandwidth of the signal acquisition path is critical for accurate characterization of the DUT behavior. This bandwidth should be at least five times wider than that of the input signal due to the spectral regrowth generated by third-, fifth-, and higher-order intermodulation products. Currently, the bandwidths of typical commercial receivers that have sufficient dynamic range for behavioral modeling and digital predistortion applications are limited to 80 MHz. Digital signal processing techniques mainly based on frequency stitching have been proposed to further increase the bandwidth of the receiver path [24], [25].

Following the acquisition of the measured input and output baseband waveforms, the data needs to be de-embedded to the DUT input and output reference planes. This includes power level adjustments but mainly time delay compensation. The propagation delay through the DUT will introduce a mismatch between the data samples used to calculate the instantaneous AM/AM and AM/PM characteristics of the DUT. This mismatch will translate into dispersion in the AM/AM and AM/PM characteristics that can be misinterpreted and considered as memory effects. Typically, the resolution needed for delay alignment is lower than the signal sampling rate. This requires

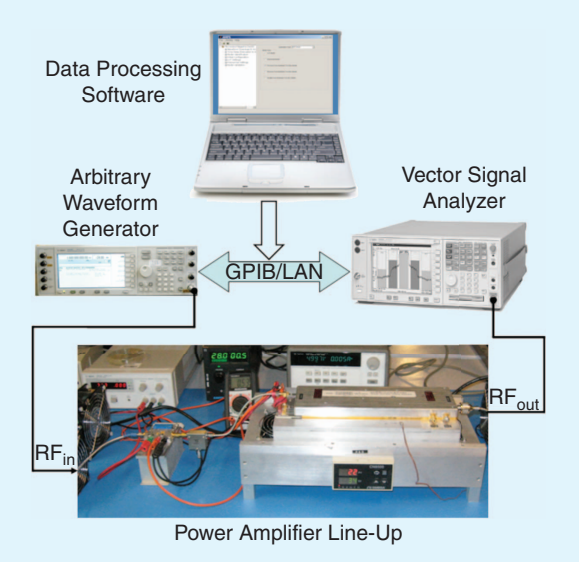

Figure 5. *Experimental setup for device-under-test characterization.*

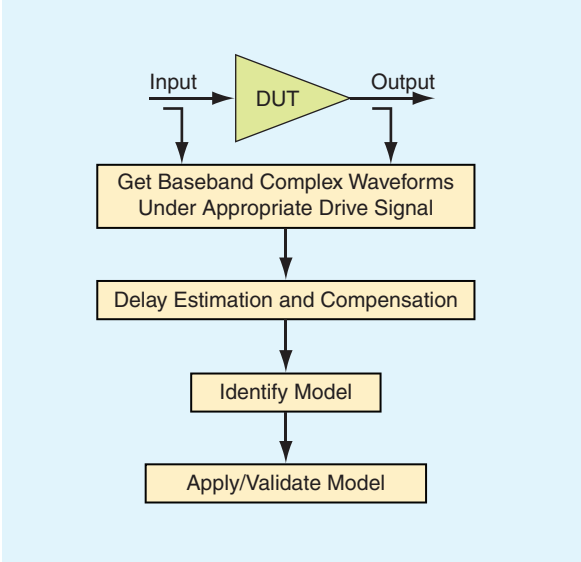

Figure 6. *Behavioral model extraction procedure: key steps from measurements to model validation.*

signal up-sampling and down-sampling during the delay estimation and compensation process [26]. The time-aligned input and output waveforms are then used to identify the behavioral model/digital predistorter of the DUT as well as its performance. These steps are summarized in Figure 6.

Even though most of the systems being considered exhibit memory effects, the look-up-table model is still used as a model for the static nonlinearity in two- and three-box-based behavioral models and digital predistorters.

Formulation

Numerous formulations have been proposed for behavioral modeling and digital predistortion of RF PAs and transmitters. These can be categorized according to several criteria, such as the inclusion or exclusion of memory effects, the number of boxes in the model, etc. Figure 7 illustrates several commonly used formulations. These include the memoryless look-up table, the nested look-up table, the Volterra, memory polynomial, envelope memory polynomial, Wiener and augmented Wiener, Hammerstein and augmented Hammerstein, and forward, reverse and parallel twin nonlinear two-box models [26]–[39]. For the description of these models, the instantaneous input and output baseband waveforms of the DUT are designated by $x_{\text{in}}(n)$ and $x_{\text{out}}(n)$, respectively.

Look-Up-Table Model

The look-up-table model is the basic behavioral model for memoryless AM/AM and AM/PM nonlinearities [27]. The complex gain of the DUT is stored in two look-up tables. The output waveform is given by:

$$x_{\text{out}}(n) = G(|x_{\text{in}}(n)|) \cdot x_{\text{in}}(n), \tag{2}$$

where $G(|x_{\text{in}}(n)|)$ is the instantaneous complex gain of the DUT.

The AM/AM and AM/PM characteristics of the DUT are derived from the raw measured data using averaging or polynomial fitting techniques.

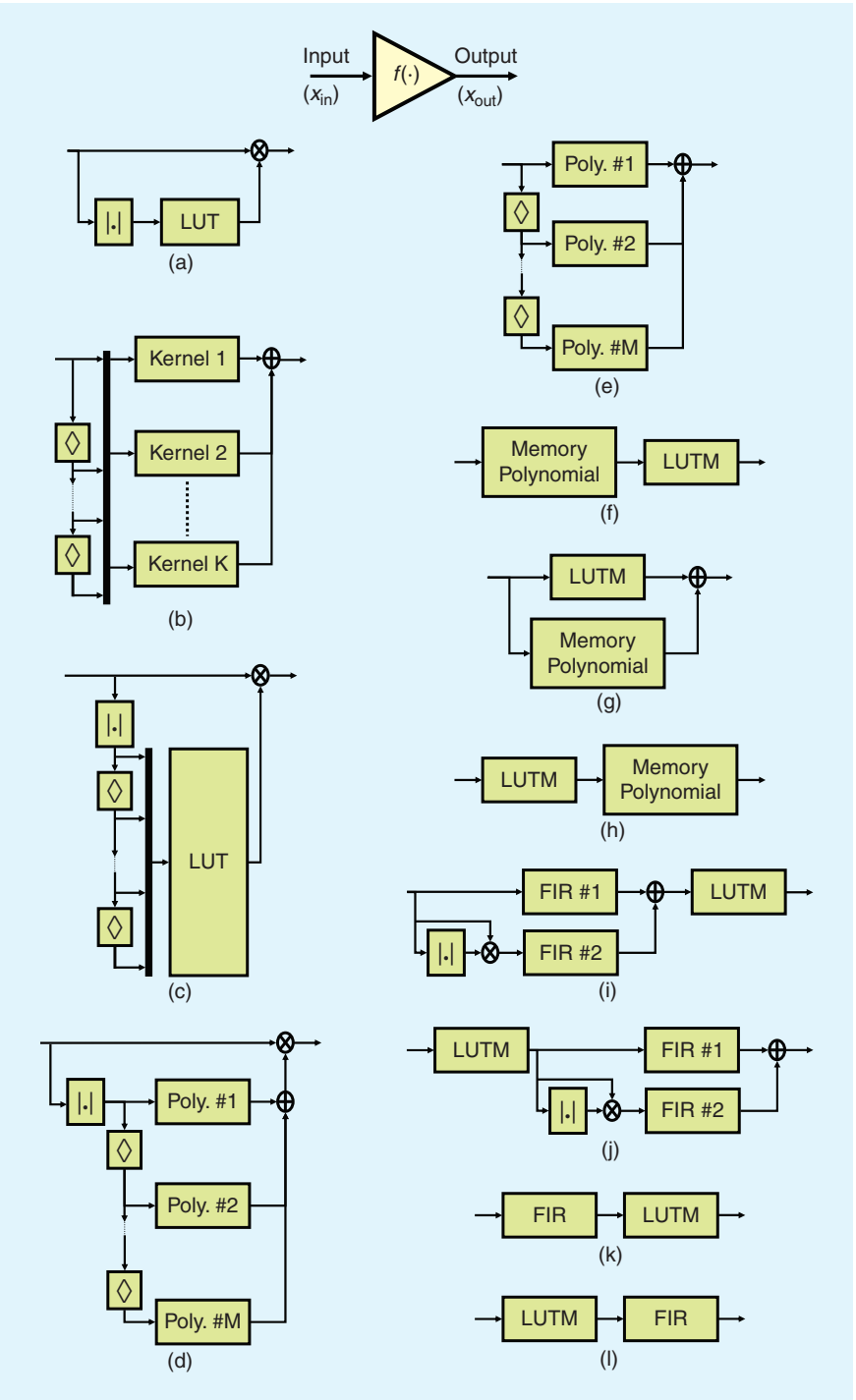

Figure 7. *State of the art structures for behavioral models and digital predistorters for power amplifiers. (a) Look-up-table (LUT) model. (b) Volterra model. (c) Nested LUT model. (d) Envelope memory polynomial (Poly) model. (e) Memory polynomial model. (f) Reverse twin nonlinear model. (g) Parallel twin nonlinear model. (h) Forward twin nonlinear model. (i) Augmented Wiener model. (j) Augmented Hammerstein model. (k) Wiener model. (l) Hammerstein model. FIR: Finite impulse response filter.*

Compensation is needed to ensure the compliance of wireless communication equipment with the mandatory linearity requirements of the communication standard being considered.

Nested Look-Up-Table Model

The nested look-up-table model was proposed to augment the conventional look-up-table-based complex-multiplier model and include memory effects [28]. For this purpose, the instantaneous gain of the DUT is a function of the actual input sample $x_{\text{in}}(n)$ and the $M-1$ preceding samples $[x_{\text{in}}(n-1), x_{\text{in}}(n-2), \ldots, x_{\text{in}}(n-M)]$ (where M is the memory depth of the DUT). Hence, the look-up table size is K^{M+1}, where K is the number of bins required for the memoryless look-up table model.

The output waveform is given by

$$x_{\text{out}}(n) = G(|X_{\text{in}}(n)|) \cdot x_{\text{in}}(n), \tag{3}$$

where $G(|X_{\text{in}}(n)|)$ is the instantaneous complex gain of the DUT, and $X_{\text{in}}(n)$ is the input vector including the present and the $M-1$ preceding samples. $X_{\text{in}}(n)$ is defined as

$$X_{\text{in}}(n) = [x_{\text{in}}(n), x_{\text{in}}(n-1), \ldots, x_{\text{in}}(n-M)]. \tag{4}$$

This model does not require any curve fitting of the measured data. The complex gain values are derived from the raw measurement after accurate time delay alignment since the memory effects information is embedded in the dispersion of the AM/AM and AM/PM characteristics.

Volterra Model

The Volterra model is the most comprehensive model for dynamic nonlinear systems. In this model, the relationship between the input and output waveforms is:

$$x_{\text{out}}(n) = \sum_{k=1}^{K} \sum_{i_1=0}^{M} \cdots \sum_{i_p=0}^{M} h_p(i_1, \cdots, i_p) \prod_{j=1}^{k} x_{\text{in}}(n-i_j), \tag{5}$$

where $h_p(i_1, \cdots, i_p)$ are the parameters of the Volterra model, K is the nonlinearity order of the model, and M is the memory depth.

The number of parameters in the conventional Volterra series increases drastically with the nonlinearity order and the memory depth. In the past, this limited the practical use of the Volterra series to weakly nonlinear systems with low-order nonlinearity. To alleviate this complexity burden, several techniques have been proposed to simplify the Volterra model. These include the pruning techniques and the dynamic reduction deviation technique [1]–[4], [29], [30]. Volterra-based models demonstrate high accuracy in modeling mildly nonlinear PAs and transmitters. For strongly nonlinear DUTs, these models may present higher complexity (in terms of number of parameters) compared to other state-of-the-art models while achieving comparable performance.

Memory Polynomial Model

The memory polynomial model is widely used for behavioral modeling and digital predistortion of PAs/transmitters exhibiting memory effects [31]–[35]. It corresponds to a reduction of the Volterra series in which only diagonal terms are kept. The output waveform of the model is

$$x_{\text{out}}(n) = \sum_{j=0}^{M} \sum_{i=1}^{N} a_{\text{ji}} \cdot x_{\text{in}}(n-j) \cdot |x_{\text{in}}(n-j)|^{i-1}, \tag{6}$$

where N and M are the nonlinearity order and the memory depth of the DUT, respectively, and a_{ji} are the model coefficients.

Several variations of the memory polynomial model have been proposed in the literature. These include the orthogonal memory polynomial model, the memory polynomial model with cross-terms also referred to as the *generalized memory polynomial model* [34], [35].

Envelope Memory Polynomial Model

The envelope memory polynomial model can be seen as a combination between the memory polynomial model and the nested look-up model [36]. The output waveform of the envelope memory polynomial model is given by

$$x_{\text{out}}(n) = \sum_{j=0}^{M} \sum_{i=1}^{N} a_{\text{ji}} \cdot x_{\text{in}}(n) \cdot |x_{\text{in}}(n-j)|^{i-1}, \tag{7}$$

where N and M are the nonlinearity order and the memory depth of the DUT, respectively, and a_{ji} are the model coefficients.

This formulation is similar to that of the memory polynomial model, except that only the magnitude information of the memory terms $[x_{\text{in}}(n-1), x_{\text{in}}(n-2), \ldots, x_{\text{in}}(n-M)]$ is required and are not their complex values. This makes this model straightforward to use for implementation in RF digital predistortion systems. As illustrated by (8), it can be seen as an implementation of the nested look-up table model that takes advantage of the compact formulation and straightforward identification of memory polynomial models.

$$G(|X_{\text{in}}(n)|) = \sum_{j=0}^{M} \sum_{i=1}^{N} a_{\text{ji}} \cdot |x_{\text{in}}(n-j)|^{i-1}. \tag{8}$$

Wiener Model

The Wiener model is a two-box model composed of a linear finite impulse response (FIR) filter followed by

a memoryless nonlinear function [26], [34], [37]. The output of this model is given by

$$x_{\text{out}}(n) = G(|x_1(n)|) \cdot x_1(n), \tag{9}$$

where $G(|x_1(n)|)$ is the memoryless instantaneous gain function implemented in the look-up table model. $x_1(n)$ designates the output of the FIR filter

$$x_1(n) = \sum_{j=0}^{M} h(j) \cdot x_{\text{in}}(n-j), \tag{10}$$

where $h(j)$ are the coefficients of the FIR filter impulse response, and M is the memory depth of the DUT.

In this model, the static nonlinear function (look-up table model) is commonly identified in the first step. Then, the input and output waveforms of the FIR filter are de-embedded and the filter coefficients identified. This approach circumvents the identification of a nonlinear system of equations.

Hammerstein Model

In the Hammerstein model, the static nonlinearity is applied upstream of the linear filter [34], [37], [38]. Thus, the output waveform is given by:

$$x_{\text{out}}(n) = \sum_{j=0}^{M} h(j) \cdot x_1(n-j) \tag{11}$$

and

$$x_1(n) = G(|x_{\text{in}}(n)|) \cdot x_{\text{in}}(n), \tag{12}$$

where $x_1(n)$, $h(j)$, and $G(|x_{\text{in}}(n)|)$ refer to the output of the first box (look-up table model), the impulse response of the FIR filter, and the instantaneous gain of the look-up table model, respectively, and M is the memory depth of the DUT.

The identification procedure of the model parameters is similar to that of the Wiener model (the look-up table model is identified first and then the filter coefficients).

Augmented Hammerstein Model

In conventional Hammerstein model, the linear filter corrects for the electrical memory effects attributed to the frequency response of the DUT around the carrier frequency but does not take into account the impedance variation of the bias circuits and harmonic loading. For this purpose, a second branch including a FIR filter applied to a second order nonlinearity was added in the augmented Hammerstein model [38]. Thus, the output waveform is now

$$x_{\text{out}}(n) = \sum_{j_1=0}^{M_1} h_1(j_1) \cdot x_1(n-j_1) + \sum_{j_2=0}^{M_2} h_2(j_2) \cdot x_1(n-j_2) \cdot |x_1(n-j_2)|, \tag{13}$$

where $h_1(j_1)$ and $h_2(j_2)$ are the impulse responses of the filters FIR #1 and FIR #2. $x_1(n)$ is the output of the look-up table box as expressed by (12). The identification of the model is done in two steps similarly to the case of the Hammerstein model.

In a conventional Hammerstein model, the linear filter corrects for the electrical memory effects attributed to the frequency response of the DUT around the carrier frequency but does not take into account the impedance variation of the bias circuits and harmonic loading.

Augmented Wiener Model

The augmented Wiener model is the dual of the augmented Hammerstein model in which the look-up table model block is applied downstream of the mildly nonlinear dynamic function [26]. The model output $x_{\text{out}}(n)$ is given by (9), yet, the intermediate signal $x_1(n)$ is now:

$$x_1(n) = \sum_{j_1=0}^{M_1} h_1(j_1) \cdot x_{\text{in}}(n-j_1) + \sum_{j_2=0}^{M_2} h_2(j_2) \cdot x_{\text{in}}(n-j_2) \cdot |x_{\text{in}}(n-j_2)|, \tag{14}$$

where $h_1(j_1)$ and $h_2(j_2)$ are the impulse responses of the filters FIR1 and FIR2 and M_1 and M_2 are memory depths of the first and second FIR filters, respectively.

Twin Nonlinear Two-Box Models

Twin nonlinear two-box models are a class of two-box models that consists of a static look-up-table-based nonlinear function and a memory polynomial dynamic nonlinear function [39]. The two boxes can be arranged in different configurations to define the forward twin nonlinear two-box model (where the look-up table model is upstream of the memory polynomial function), the reverse twin nonlinear two-box model (in which the memory polynomial function is applied prior to the look-up table model), and the parallel twin nonlinear two-box model (where both the look-up table and memory polynomial functions are applied to the input signal).

For each of the three twin nonlinear two-box models, the output waveform can be derived as a function of the input waveform through a proper combination of (2) and (6).

All the aforementioned structures are being used for the modeling of PAs and transmitters exhibiting dynamic nonlinear behavior. Through these models, one can distinguish three trends that are leading to new models.

- *Wiener- and Hammerstein-based models:* Augmenting the original Wiener and Hammerstein model

to include mild nonlinearities in the dynamic box and also, in some cases, combining the Wiener and Hammerstein models in three-box structures such as Wiener-Hammerstein and Hammerstein-Wiener models [34], [40], [41].

- *Volterra based models:* Pruning and reducing the original Volterra-series-based model to keep only the terms that have noticeable impact on the DUT behavior.
- *Memory polynomial based models:* Augmenting the original memory polynomial model to include additional specific terms or reducing/reformulating the memory polynomial model to reduce its complexity and/or enhance its performance.

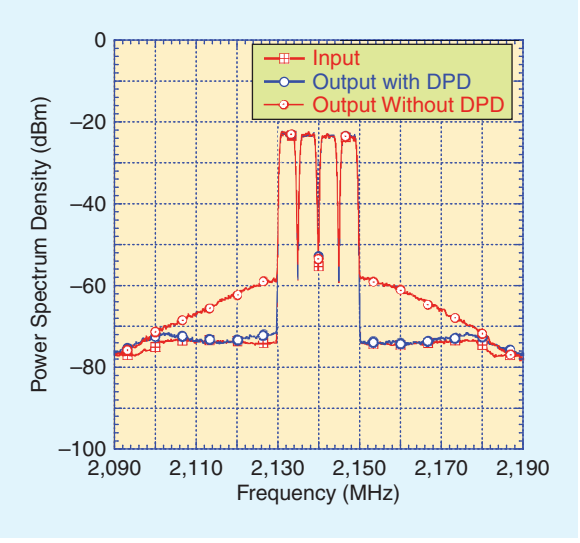

Figure 8. *Digital predistortion performance on a GaN-based Doherty power amplifier operating at full power (output power backoff equal to the signal's peak-to-average power ratio) shows quasi-perfect cancellation of the distortion.*

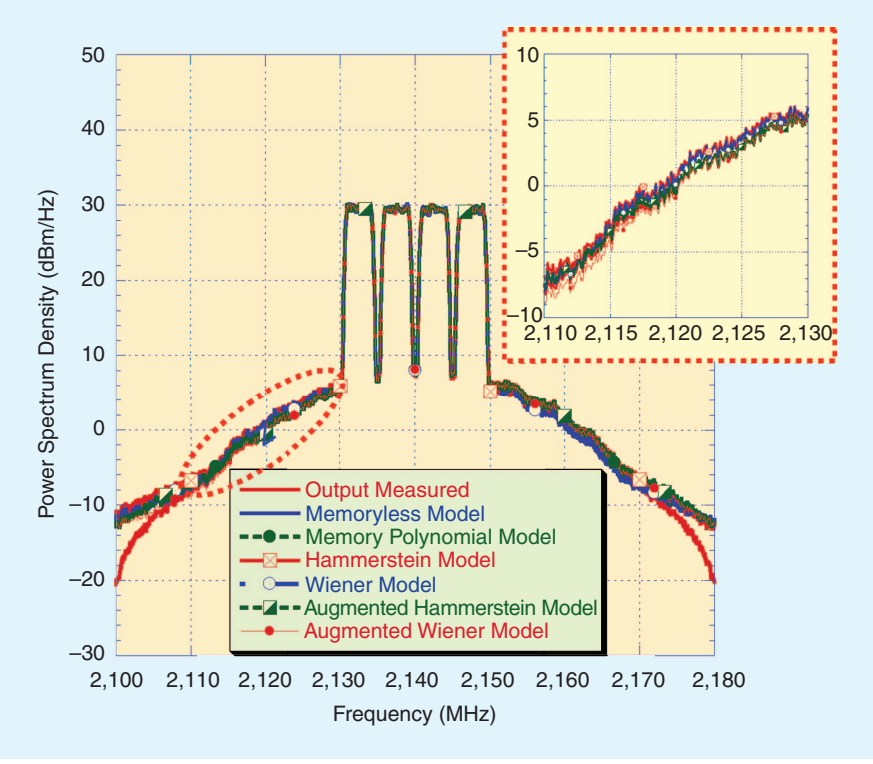

Figure 9. *Measured and estimated output spectra before memoryless postcompensation. All models used demonstrate similar performance. (From [20].)*

Comparison and Discussion of Model Structures

It is essential to have access to accurate metrics for the performance assessment of the different structures reported in Figure 7. For digital predistortion applications, the performance evaluation is straightforward and can be done based on the linearity metrics either in the frequency domain using the ACPR or in the time domain based on error vector magnitude (EVM). Figure 8 presents the measured spectrum at the output of a linearized GaN based Doherty PA following memory polynomial-based digital predistortion. This figure clearly shows the perfect cancellation of the nonlinearities exhibited by the DUT because the output of the linearized PA is quasi-identical to its scaled input signal.

Several metrics have been proposed for assessing the performance of behavioral models in either the frequency or the time domain. These include the well-established normalized mean square error (NMSE), the memory effects modeling ratio (MEMR), the adjacent channel error power ratio (ACEPR), and the weighted error-to-signal power ratio (WESPR) [20], [30], [42]–[46]. The NMSE and, equivalently, MEMR are dominated by the in-band error. Conversely, ACEPR and WESPR metrics provide a better insight of the model performance in the adjacent channel (out-of-band error). Moreover, in out-of-band error, the contribution of memory effects is frequently buried under that of the static nonlinearity in highly nonlinear devices such as Doherty PAs.

For accurate evaluation of the modeling of the memory effects, static nonlinearity cancellation techniques have been proposed. This can be achieved either through the use of a memoryless digital predistorter upstream of the DUT and its model or the use of a memoryless digital postdistorter at the output of the DUT and its model [20], [47]. Further

model validation can be performed by applying digital predistortion or, equivalently, digital postcompensation for both the static and dynamic distortion of the DUT and comparing the resulting signals at the output of the linearized DUT and its linearized model. The accuracies of predistortion and postcompensation in assessing behavior model performance are equivalent. However, the use of postcompensation is more suitable for behavioral modeling applications since it requires fewer measurements compared to the use of a predistorter. In practice, to apply predistortion, an access to the DUT and predistortion setup is required. However, the use of postcompensation only requires the availability of the measurement data and not the measurement setup since all the processing can be done using simulation software.

Figure 9 presents measured spectra at the output of a 300-W high-efficiency lateral double-diffused metal-oxide–semiconductor field-effect transitor (LDMOS) based Doherty PA as well as the estimated spectra using several models among those presented in Figure 7. According to these results, all the models, even the memoryless look-up table, appear to accurately predict the behavior of the DUT. However, as shown in Figure 10, there is a clear difference in the behavior of the same models and their ability to predict the output waveform of the DUT after static nonlinearity cancellation using a memoryless postcompensation technique. Thus, the use of a memoryless postcompensation technique enables accurate assessment of the model performance in mimicking the DUT's memory effects [20].

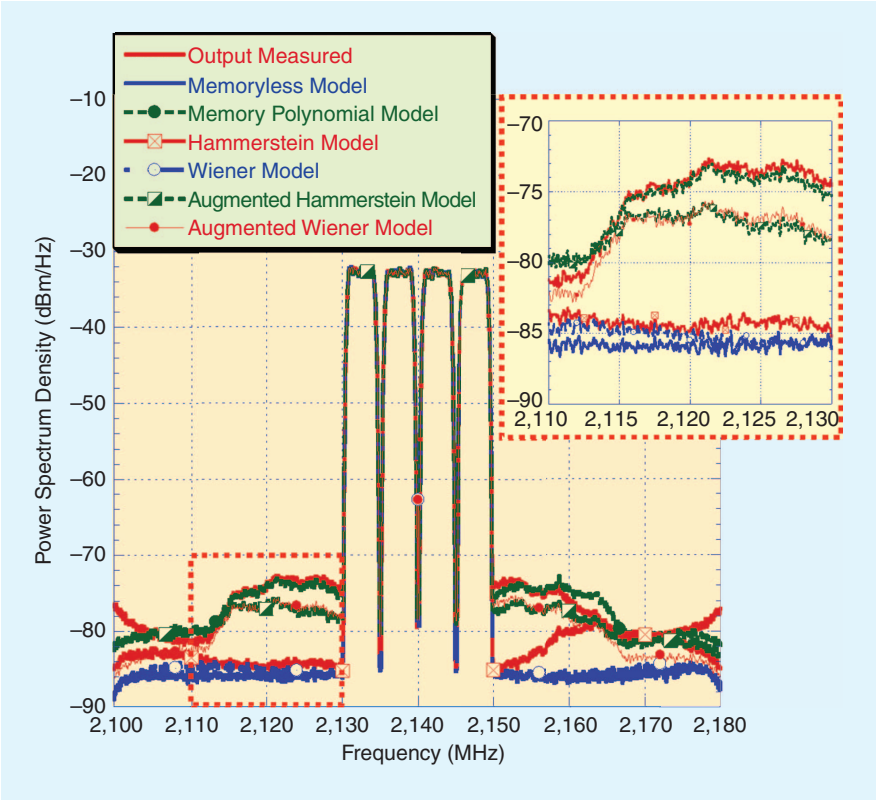

Figure 10. *Measured and estimated output spectra after memoryless postcompensation. The differentiation between the performance of the various models is possible following memoryless-postcompensation. (From [20].)*

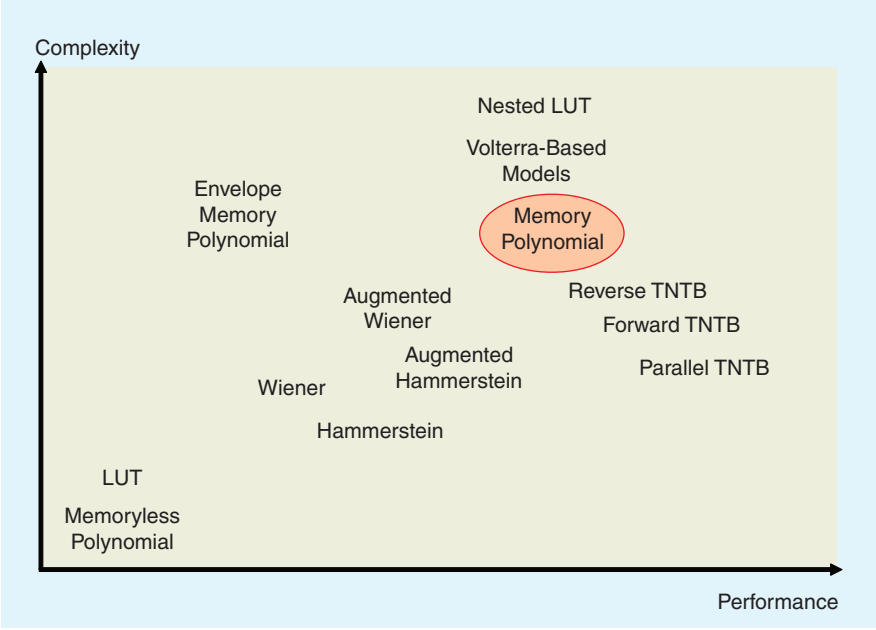

Figure 11. *Comparison between state-of-the-art structures for power amplifier behavioral models and digital predistorters. LUT: Look-up table. TNTB: Twin nonlinear two-box models.*

According to the previous discussion, none of the figures of merit that have been proposed so far presents a reliable metric for comprehensive model performance assessment that takes into account the in-band and out-of-band errors including the static nonlinear behavior and the memory effects.

Compared to the augmented Wiener and Hammerstein models, the twin nonlinear two-box class achieves better performance at the expense of a slightly higher complexity.

This makes the comparison of different structures in behavioral modeling context even more delicate. Such comparison needs to take into account their performance as well as their complexity. So far, there is no reported study comparing head to head all the previously described models illustrated in Figure 7. Yet, the results are likely to be device dependant, especially for cascaded two-box models, which attempt to separate the behavior of the DUT into a static nonlinear behavior and a linear/mildly nonlinear dynamic function.

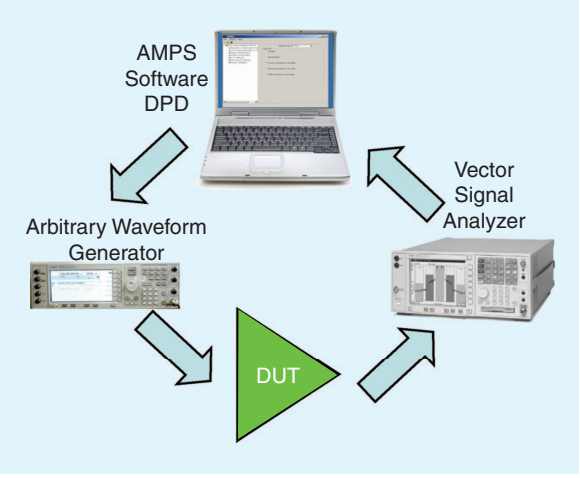

Figure 12. *Automated closed-loop digital predistorter using commercial instruments.*

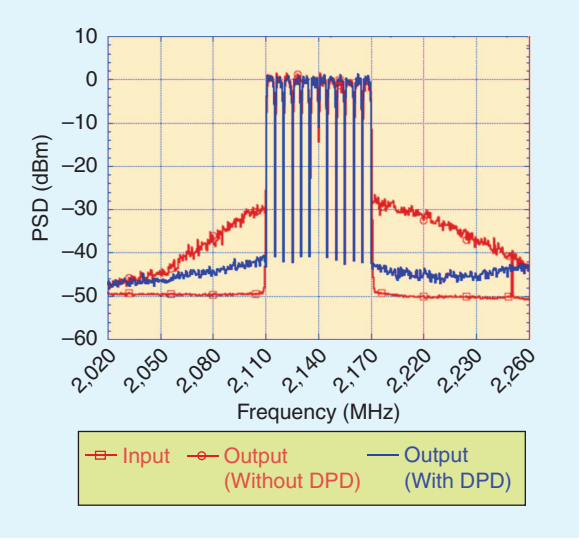

Figure 13. *Digital predistortion performance on a GaN-based Doherty power amplifier operating at full power (output power backoff equal to the signal's peak-to-average power ratio). Significant spectral regrowth cancellation is observed for a 12-carrier WCDMA input signal.*

Based on a compilation of the results presented in the open literature in behavioral modeling and digital predistortion contexts, an attempt to fairly qualitatively compare the different model structures is illustrated in Figure 11 [4], [26]–[39]. This figure presents a general trend for relative performance and complexity built around the comparison of the considered models with the standard memory polynomial model. The nonlinear system considered herein is a typical, strongly nonlinear system such as Doherty PA that requires at least tenth-order static nonlinearity and exhibits at least a third-order memory depth.

In Figure 11, it is important to notice that all models except the memoryless ones are clustered around the memory polynomial and the Volterra model. These models outperform most of the models except the twin nonlinear two-box class. This is mainly due to the identification procedure involved in the twin nonlinear two-box models, which decouples the contribution of the static and dynamic behaviors. In the memory polynomial model, the identification of both behaviors is performed simultaneously. The error function that is minimized for the model parameter identification is typically dominated by the static contribution. Compared to the augmented Wiener and Hammerstein models, the twin nonlinear two-box class achieves better performance since it models the dynamic behavior by a memory polynomial function while this same behavior is modeled using a first- and-second order nonlinearity in the case of augmented Wiener and Hammerstein models. This performance improvement is achieved at the expense of a slightly higher complexity.

Within the twin nonlinear two-box class, the forward and reverse models have similar performance and are both less accurate than the parallel twin nonlinear two-box model. This can be attributed to the fact that the structure of the first two models inherently forces a cascade of the static and dynamic behaviors of the DUT. However, the parallel twin nonlinear two-box model is more generic since it does not assume such a cascaded structure. There is no clear comparison between the performance of memory polynomial models to those of Volterra based models. Because memory polynomials are simplified versions of Volterra models, it is expected that both will lead to comparable performance.

The complexity of two-box based models (including Wiener, Hammerstein, and their augmented version, as well as twin nonlinear two-box class) is lower than that of the memory polynomial

model. Even though these two-box models involve an extra de-embedding step, they typically require a much lower number of parameters. Conversely, Volterra-based models can lead to higher complexity. Finally, the complexity and performance of the nested look-up table model can be controlled through the choice of the number of bins in the look-up table. Often, this model has high complexity for accuracy comparable to the memory polynomial model.

Automated Closed-Loop Digital Predistortion

It is crucial for PA designers to have access to powerful and flexible digital predistortion platforms to assess the linearizability of PA prototypes. Figure 12 presents a fully automated closed-loop software-based digital predistortion solution. This digital predistortion software, developed at the University of Calgary, can implement most of the structures presented in Figure 7. It seamlessly operates at any carrier frequency and bandwidth that can be handled by the commercial instruments used to generate and acquire the DUT's input and output RF signals. This allows straightforward linearization of PAs for any input signal. Performance with up to 60 MHz bandwidth can be achieved, as illustrated in Figure 13, which presents the measured spectra at the output of a GaN based Doherty PA before and after memory-polynomial-based digital predistortion for a 12-carrier 60-MHz wide WCDMA signal.

Conclusion

In this article, a thorough overview of behavioral modeling and predistortion of dynamic nonlinearities in RF PAs and transmitters was presented.

The sensitivity of the DUT behavior to the characteristics of the stimulus was reviewed to ensure appropriate conditions for accurate observation. Nearly all state-of-the-art behavioral models were described and their relative performance and complexity discussed.

Similarities and specifics of behavioral modeling and digital predistortion were presented. Thereby, digital predistortion can be seen as a behavioral modeling problem for which performance assessment is much more straightforward. For DUT behavioral modeling, there is no comprehensive metric that allows the model performance evaluation while taking into account the model accuracy in predicting all the three components of the DUT behavior (in-band distortion, static nonlinearity and memory effects). Finally, a software digital predistortion solution that enables closed-loop wideband linearization was briefly presented with excellent linearization capabilities when amplifying a 12-carrier 60-MHZ wide WCDMA signal.

It is crucial for PA designers to have access to powerful and flexible digital predistortion platforms to assess the linearizability of PA prototypes.

Acknowledgment

This work was supported by the Alberta Informatics Circle of Research Excellence (iCORE), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Canada Research Chair (CRC) Program.

References

[1] D. Schreurs, M. O'Droma, A. A. Goacher, and M. Gadringer, *RF Power Amplifier Behavioral Modeling*. New York, NY: Cambridge Univ. Press, 2008.

[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. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 223–227, Jan. 1997.

[4] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Open-loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 7, pp. 1524–1534, July 2008.

[5] K. J. Muhonen, M. Kavehrad, and R. Krishnamoorthy, "Look-up table techniques for adaptive digital predistortion: A development and comparison," *IEEE Trans. Veh. Technol.*, vol. 49, no. 5, pp. 2118–2127, May 2006.

[6] O. Hammi and F. M. Ghannouchi, "Power alignment of digital predistorters for power amplifiers linearity optimization," *IEEE Trans. Broadcast.*, vol. 55, no. 1, pp. 109–114, Mar. 2009.

[7] A. A. Moulthrop, C. J. Clark, C. P. Silva, and M. S. Muha, "A dynamic AM/AM and AM/PM measurement technique," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 1997, vol. 3, pp. 1455–1458.

[8] H. Ku, M. D. Mckinley, and J. S. Kenney, "Extraction of accurate behavioral models for power amplifiers with memory effects using two-tone measurements," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2002, pp. 139–142.

[9] Y. Yang, J. Yi, J. Nam, B. Kim, and M. Park, "Measurement of two-transfer characteristics of high-power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 49, no. 3, pp. 568–571, Mar. 2001.

[10] D. Schreurs, M. Myslinski, and K. Remley, "RF behavioral modeling from multisine measurements: Influence of excitation type," in *Proc. 33rd European Microwave Conf.*, Oct. 2003, pp. 1011–1014.

[11] K. Remley, "Multisine excitation for ACPR measurements," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2003, pp. 2141–2144.

[12] J. C. Pedro and N. B. Carvalho, "Designing multisine excitations for nonlinear model testing," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 1, pp. 45–54, Jan. 2005.

[13] O. Anderson, P. Malmlof, and D. Wisell, "Nonlinear characterization of multiple carrier power amplifiers," in *56th ARFTG Conf. Dig.*, Nov. 2000, pp. 111–118.

[14] S. Boumaiza, M. Helaoui, O. Hammi, T. Liu, and F. M. Ghannouchi, "Systematic and adaptive characterization approach for behavior

modeling and correction of dynamic nonlinear transmitters," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 6, pp. 2203–2211, Dec. 2007.

[15] J. C. Pedro and N. B. Carvalho, "On the use of multitone techniques for assessing RF components' intermodulation distortion," *IEEE Trans. Microw. Theory Tech.*, vol. 47, no. 12, pp. 2393–2402, Dec. 1999.

[16] N. Boulejfen, A. Harguem, and F. M. Ghannouchi, "New closed-form expressions for the prediction of multitone intermodulation distortion in fifth-order nonlinear RF circuits / systems," *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 1, pp. 121–132, Jan. 2004.

[17] F. P. Hart and M. B. Steer, "Modeling of nonlinear response of multitones with uncorrelated phase," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 10, pp. 2147–2156, Oct. 2007.

[18] O. Hammi, S. Carichner, B. Vassilakis, and F. M. Ghannouchi, "Synergetic crest factor reduction and baseband digital predistortion for adaptive 3G Doherty power amplifier linearizer design," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 11, pp. 2602–2608, Nov. 2008.

[19] O. Hammi, S. Carichner, B. Vassilakis, and F. M. Ghannouchi, "Novel approach for static nonlinear behavior identification in RF power amplifiers exhibiting memory effects," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2008, pp. 1521–1524.

[20] O. Hammi, S. Carichner, B. Vassilakis, and F. M. Ghannouchi, "Power amplifiers' model assessment and memory effects intensity quantification using memoryless post-compensation technique," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 12, pp. 3170–3179, Dec. 2008.

[21] J. P. Martins, P. M. Cabral, N. B. Carvalho, and J. C. Pedro, "A metric for the quantification of memory effects in power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4432–4439, Dec. 2006.

[22] H. C. Ku, M. D. 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.

[23] P. J. Draxler, A. Zhu, J. J. Yan, P. Kolinko, D. F. Kimball, and P. M. Asbeck, "Quantifying distortion of RF power amplifiers for estimation of predistorter performance," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2008, pp. 931–934.

[24] D. Widell, D. Ronnow, and P. Handel, "A technique to extend the bandwidth of a power amplifier test-bed," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 4, pp. 1488–1494, Aug. 2007.

[25] P. N. Landin, C. Nader, N. Bjorsell, M. Isaksson, D. Wisell, P. Handel, O. Andresen, and N. Keskitalo, "Wideband characterization of power amplifiers using undersampling," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2009, pp. 1365–1368.

[26] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Deembedding static nonlinearities and accurately identifying and modeling memory effects in wide-band RF transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 11, pp. 3578–3587, Nov. 2005.

[27] J. K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 374–382, Nov. 1990.

[28] O. Hammi, F. M. Ghannouchi, S. Boumaiza, and B. Vassilakis, "A data-based nested LUT model for RF power amplifiers exhibiting memory effects," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 10, pp. 712–714, Oct. 2007.

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

[30] A. Zhu, J. Pedro, and T. Cunha, "Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge," *IEEE Trans. Microw. Theory Tech.*, vol. 55, no. 5, pp. 813–821, May 2007.

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

[32] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[33] R. N. Braithwaite, "Wide bandwidth adaptive digital predistortion of power amplifiers using reduced order memory correction," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2008, pp. 1517–1520.

[34] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, pp. 3852–3860, Oct. 2006.

[35] R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 1468–1479, Sept. 2004.

[36] O. Hammi, F. M. Ghannouchi, and B. Vassilakis, "A compact envelope-memory polynomial for RF transmitters modeling with application to baseband and RF-digital predistortion," *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 5, pp. 359–361, May 2008.

[37] P. Gilabert, G. Montoro, and E. Bertran, "On the Wiener and Hammerstein models for power amplifier predistortion," in *Microwave Conf. Proc.*, APMC 2005, Asia-Pacific, Dec. 2005.

[38] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Augmented Hammerstein predistorter for linearization of broad-band wireless transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1340–1349, Apr. 2006.

[39] O. Hammi and F. M. Ghannouchi, "Twin nonlinear two-box models for power amplifiers and transmitters exhibiting memory effects with application to digital predistortion," *IEEE Microw. Wireless Compon. Lett.*, vol. 19, no. 8, pp. 530–532, Aug. 2009.

[40] P. Crama and Y. Rolain, "Broadband measurement and identification of a Wiener-Hammerstein model for an RF amplifier," in *60th ARFTG Conf. Dig.*, Dec. 2002, pp. 49–57.

[41] M. Sano and L. Sun, "Identification of Hammerstein-Wiener system with application to compensation for nonlinear distortion," in *41st SICE Annu. Conf. Proc.*, Aug. 2002, pp. 1521–1526.

[42] P. Landin, M. Isaksson, and P. Handel, "Comparison of evaluation criteria for power amplifier behavioral modeling," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2008, pp. 1441–1444.

[43] P. L. Gilabert, D. D. Silveira, G. Montoro, M. E. Gardinger, and E. Bertran, "Heuristic algorithms for power amplifier behavioral modeling," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 10, pp. 715–717, Oct. 2007.

[44] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 1, pp. 348–359, Jan. 2006.

[45] H. Ku and J. S. Kenney, "Behavioral modeling of nonlinear RF power amplifiers considering memory effects," *IEEE Trans. Microw. Theory Tech.*, vol. 51, no. 12, pp. 2495–2504, Dec. 2003.

[46] D. Wisell, M. Isaksson, and N. Keskitalo, "A general evaluation criteria for behavioral power amplifier modeling," in *69th ARFTG Conf. Dig.*, June 2007, pp. 251–255.

[47] T. Liu, S. Boumaiza, A. B. Sesay, and F. M. Ghannouchi, "Quantitative measurements of memory effects in wideband RF power amplifiers driven by modulated signals," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 1, pp. 79–81, Jan. 2007.