Digital Predistortion for Envelope-Tracking Power Amplifiers Using Decomposed Piecewise Volterra Series

Anding Zhu, Member, IEEE, Paul J. Draxler, Member, IEEE, Chin Hsia, Thomas J. Brazil, Fellow, IEEE, Donald F. Kimball, Member, IEEE, and Peter M. Asbeck, Fellow, IEEE

Abstract—Due to dynamic changes of supply voltage, envelope-tracking (ET) power amplifiers (PAs) exhibit very distinct characteristics in different power regions. It is very difficult to compensate the distortion induced by these amplifiers by employing conventional digital predistortion techniques. In this paper, by introducing a new piecewise Volterra model based on a vector threshold decomposition technique, we first set several thresholds in the input power level according to the PA characteristics, and decompose the input complex envelope signal into several sub-signals by using these thresholds. We then process each sub-signal separately by employing the dynamic deviation reduction-based Volterra series, and finally recombine them together to produce the predistorted output. Experimental results show that by using this new decomposed piecewise digital predistorter model, the distinct characteristics of the ET system at different signal power levels can be accurately modeled, and thus, the distortion, including both static nonlinearities and memory effects, caused by the amplifier nonlinear behavior can be effectively compensated.

Index Terms—Behavioral modeling, envelope tracking (ET), linearization, power amplifier (PA), predistorter, Volterra series.

I. INTRODUCTION

In modern wireless communication systems, nonconstant envelope modulation techniques are normally employed. The traditional approach to linearly amplifying these nonconstant envelope signals is to “backoff” the output power of RF power amplifiers (PAs) until the distortion level is within acceptable limits. Unfortunately, this results in low power efficiency since these modulated signals often have a high peak-to-average power ratio (PAPR) [1].

Envelope tracking (ET) is one of the most promising techniques for high efficiency operation for nonconstant envelope signals [2]. By superimposing the envelope signal at the drain voltage, the RF transistor in the PA can be operated continuously in the compression regime over a wide range of power levels, which significantly increases power efficiency of the PA. However, the ET system has inherent nonlinearities associated with the gain and phase variations as the drain voltage changes. As standard RF PAs, for wideband and high-power operation, memory effects will also become apparent in the ET system. In order to improve the system performance, it is necessary to employ digital predistortion techniques to remove nonlinear distortion and memory effects [3].

Previously, we proposed an open-loop digital predistorter (DPD) for RF PAs with a fixed supply voltage in [4]. This DPD approach is derived from the dynamic deviation reduction-based Volterra series [5] that allows simultaneous compensation for nonlinear distortion and memory effects with a small number of parameters. Based on the pth-order post-inverse theory [6], the parameters of this DPD can be directly estimated from the measured input and output of the PA with a simple offline characterization process. This eliminates the real-time closed-loop adaptation requirement, and removes the necessity to implement the parameter-estimation algorithms in real digital circuits, which significantly reduces system complexity and implementation cost. However, this type of model cannot be applied directly to the ET system. Since the amplifier in the ET system exhibits very distinct characteristics at the different power levels, it is very difficult to characterize the behavior of the ET PA by using a single dynamic Volterra function for the entire range of the input signal level.

In this paper, we propose an extension to the Volterra model employed in [4] so that the DPD approach proposed in [4] can be used to linearize the ET system. This is achieved by introducing a new decomposed piecewise Volterra model based on a vector threshold decomposition technique. In this approach, we first set several thresholds in the input power level according to the characteristics of the PA. We then decompose the input complex envelope signal into several sub-signals by using these thresholds, and continue to process each sub-signal separately by employing a low-order dynamic deviation reduction-based Volterra series, and finally recombine them together to produce the predistorted output. By employing this new piecewise type of DPD model, the distinct characteristics of the ET system at different signal power levels can be accurately modeled, and thus, the distortion caused by the nonlinear behavior of the ET system can be effectively compensated. As in the model in [4]...
and [5], the output of this new model is also linear in relation to all of its parameters so that the model can be easily characterized by using the same linear estimation algorithms as presented in [4].

This paper is organized as follows. In Section II, we briefly introduce the ET system. Section III presents some conventional approaches to linearizing the ET system. Section IV introduces the vector threshold decomposition technique and the new digital predistortion model structure. Parameter-extraction procedures and system implementation are also discussed in this section. The experimental results are given in Section V. A conclusion is then presented in Section VI.

II. ET SYSTEM

A block diagram of an ET system is shown in Fig. 1. The envelope signal is first detected from the original input signal and then amplified by an envelope amplifier, providing a dynamic drain (or collector) voltage to the RF transistor in the main PA. At the same time, the RF signal is amplified by the main RF PA. The delay line is to compensate the misalignment between the envelope and RF signal. In modern ET systems, the envelop signal can be directly generated from the baseband in-phase/quadrature (I/Q) waveforms [1].

In the ET system, the gate (or base) of the RF transistor in the main PA is normally biased in the class AB or class C mode, but the drain (or collector) voltage is dynamically controlled by the envelope amplifier and is changing in proportion to the envelope of the RF input. This tracking process can be continued down to a selected point on the low-voltage side, and up to a point where the RF swing reaches breakdown level. Due to full rail-to-rail voltage swing, in this system, the RF PA is continuously operated in or near saturation for all envelope signal levels, and therefore, high power efficiency can be obtained [1].

Although the output of the ET PA can be kept linear with the input over a wide range, there are inherent nonlinearities in the ET system [7]–[11]. Due to dynamic changes of the supply voltage, the ET PA exploits very distinct characteristics in different power regions. This nonlinear behavior is very different from what we normally see in the typical linear amplifiers with a fixed supply voltage. For instance, Fig. 2 shows an example of gain and phase characteristics of a gallium nitride (GaN) ET PA excited by a wideband code division multiple access (WCDMA) signal, where we can see that large gain reduction occurs as the input signal (and Vdd) is changed to the lower values, which leads to gain expansion. While the power is increased to the higher level, the gain is compressed again. The phase remains flat at low and medium power, but it changes rapidly in high power. This behavior is device dependent, as will become evident in Section V, where very different characteristics are found for a silicon laterally diffused metal oxide semiconductor (Si LDMOS) PA.

In the ET system, to avoid the Vdd voltage dropping to zero, detroughing is normally employed [8]. As a result, the drain voltage may not replicate the signal envelope with great accuracy, which may introduce distortion to the system, especially at low power levels. Some memory effects from the bias networks may be reduced in the ET system compared to the fixed-Vdd PAs because the drain voltage dynamically changes with the amplitude of the envelope. For wideband and high-power operation, memory effects still become apparent. For example, thermal transients and trapping effects can introduce significant memory effects to the system. Furthermore, due to limited bandwidth of the envelope amplifier, frequency cutoff of the ET signal also causes memory effects. As shown in Fig. 2, memory effects may appear differently in different power regions. These nonlinear distortion and memory effects can significantly reduce the overall output signal quality.
III. LINEARIZING ET SYSTEM

To remove nonlinear distortion and improve system performance, DPDs are often employed in ET systems. In recent decades, various DPD systems have been proposed [3] with the main differences usually lying either in the implementation of predistortion architectures or in parameter-extraction algorithms. There are two distinct methods to implement DPD algorithms, which are: 1) lookup tables (LUTs) and 2) analytical functions.

The LUT-based DPD has been widely used because it is relatively simple and easily implemented. Moreover, it can be fitted to any type of characteristics, including that of the ET. However, while wireless systems are migrating to wideband operations, the bandwidth and dynamic range of the signals to be transmitted are expanded dramatically. In order to cover all of the operating regions, the LUT requires a large memory to store sufficient data, which significantly increases the system complexity. Furthermore, it is very difficult to embed compensation for “memory effects” into the LUT.

The analytical function-based DPDs normally require less memory storage than the LUTs, and can operate with a small number of pre-decided parameters. Additionally, memory effects can be easily built into these functions. A number of analytical function-based memory DPDs have been developed and some of them have been widely used, especially the polynomial-like models, e.g., memory polynomials [12], [13], Hammerstein and Wiener models [14], and various simplified Volterra-series-based models [15]–[18].

However, it has been found exceedingly difficult to implement these single function-based models in the ET system [8]–[11]. This is because, in this system, the characteristics of the PA are very different at different power regions. For instance, to linearize the PA shown in Fig. 2, the AM/AM and AM/PM characteristics of the DPD would look like the plots in Fig. 3, where both gain expansion and gain compression occur, and amplitude and phase change inconsistently in different regions. To fit this distinct behavior over the entire power levels, a very high degree of nonlinearity order is required if a single polynomial-like model (e.g., a Volterra model) is employed. High-order nonlinear terms lead to very poor extrapolation of the model, which sometimes dramatically degrades system performance.

One possible solution to model these distinct nonlinearities in the ET system is piecewise curve fitting, whereby we can divide the nonlinear curves into several segments based on the input power level, and then fit each segment separately using different functions [19]. For example, the output of the DPD ($\tilde{u}(n)$) can be obtained from

$$\tilde{u}(n) = \begin{cases} 
F_1[\bar{x}(n)] & 0 \leq |\bar{x}(n)| \leq \alpha_1 \\
F_2[\bar{x}(n)] & \alpha_1 < |\bar{x}(n)| \leq \alpha_2 \\
\vdots & \\
F_N[\bar{x}(n)] & \alpha_{N-1} < |\bar{x}(n)| \leq \alpha_N 
\end{cases} \tag{1}$$

where $\bar{x}(n)$ is the complex input I/Q signal, and $|\bar{x}(n)|$ returns its magnitude value. $F_i[\cdot]$ is the nonlinear transfer function for the signal whose magnitude falls in the interval $(\alpha_{i-1}, \alpha_i]$. Typically, the fitting functions $F_i[\cdot]$ can be polynomials, splines, or other similar functions, and the parameters of these functions in different segments can be identified separately.

This polynomial curve fitting has been widely employed in modeling many electronic circuits [19], [20]. An example of the implementation of this approach is shown in Fig. 4. The
time-domain samples are first distributed to different branches according to their power levels. These samples are then processed separately using different functions and finally sent to the output. Implementing this system is straightforward. However, there are several problems in using this approach to linearize the ET PAs. Firstly, since input samples in different segments are processed by different functions, we must guarantee the output values from two adjacent segments at the joint points are continuous, e.g., \( F_i[\alpha] \) must be equal to \( F_{i+1}[\alpha] \), otherwise discontinuities will occur. To minimize these discontinuities, certain additional constraints must be imposed on the piecewise functions, e.g., having a certain number of continuous derivatives, which increases model complexity and complicates model extraction. Secondly, in this piecewise curve fitting, the sub-samples are processed in series in time. It is very difficult to embed memory effects into the model since previous inputs and current sample could fall in different intervals and processed by different functions.

IV. DECOMPOSED PIECEWISE VOLterra SERIES-BASED DPD

To cope with discontinuity and memory effects, this paper proposes a new approach to linearizing the ET amplifier in which we first decompose the input complex envelope signal into several sub-signals by using a vector threshold decomposition technique, and then process each sub-signal separately using dynamic deviation reduction-based Volterra series at different power levels, and finally recombine them together to produce the predistorted output. The concept is presented in more detail below.

A. Vector Threshold Decomposition

Threshold decomposition was developed for real-valued signals in [21] and [22] by Heredia and Arce. In this section, we extend this concept to the discrete complex envelope domain so that it can be used for building digital predistortion functions for the ET system in baseband. We call this technique vector threshold decomposition.

As mentioned earlier, in the ET system, the PA behavior strongly depends on the input power level. Equivalently, in the complex baseband, the characteristics of the PA depend on the magnitude of the input I/Q waveform. To effectively model this nonlinear behavior, we define a set of thresholds

\[
\tau = \{\lambda_1, \lambda_2, \ldots, \lambda_S\}
\]  

(2)

where \( \lambda_s \) is the magnitude level of the input envelope and \( \lambda_1 < \lambda_2 < \cdots < \lambda_S \), and \( S \) is the total number of thresholds. Since the original input signal \( \tilde{x}(n) \) is a complex-valued I/Q signal, these thresholds are not a set of single real values, but a set of circles on the complex I/Q plane. In other words, \( \lambda_s \) represents the radius of the \( s \)th threshold circle on the constellation plane, as shown in Fig. 5.

By using these threshold circles, the signal space is divided into several interval zones, and the original input signal can then be decomposed into several sub-signals located in the corresponding interval zones. For instance, the \( s \)th sub-signal in zone \( s \) can be obtained from

\[
\tilde{x}_s(n) = \left\{ \begin{array}{ll}
0, & |\tilde{x}(n)| \leq \lambda_{s-1} \\
[|\tilde{x}(n)| - \lambda_{s-1}] e^{j\varphi}, & \lambda_{s-1} < |\tilde{x}(n)| \leq \lambda_s \\
\lambda_s - \lambda_{s-1}, & |\tilde{x}(n)| > \lambda_s
\end{array} \right.
\]  

(3)

where \( |\tilde{x}(n)| \) returns the magnitude value of \( \tilde{x}(n) \) and \( \varphi \) is the phase of \( \tilde{x}(n) \). Here we assume \( \lambda_0 = 0 \) and \( 1 \leq s \leq S + 1 \).

From (3), we can see that the sub-signals are only divided by the magnitude levels; the phases of all sub-signals are still the same as those of the original signal. As an example, consider a set of two thresholds, i.e., \( \lambda_1 = 0.5, \lambda_2 = 1,4 \), which partition the signal space into the three interval zones \([0, 0.5], (0.5, 1.4], \) and \((1.4, \infty)\). A particular signal value such as \( \tilde{x}(1) = \sqrt{2} + \sqrt{2}j \) can be decomposed into the three sub-signals \( \tilde{x}_1(1) = 0.5\angle 45^\circ, \tilde{x}_2(1) = 0.9\angle 45^\circ, \) and \( \tilde{x}_3(1) = 0.6\angle 45^\circ \). We can see that the magnitude of the first component \( \tilde{x}_1(1) \) is equal to \( \lambda_1 \), and the magnitude of the second component \( \tilde{x}_2(1) \) is equal to the radius differences between zone 2 and zone 1, i.e., \( \lambda_2 - \lambda_1 \), and finally the last component \( \tilde{x}_3(1) \) is obtained from the subtraction of the original magnitude value and the threshold \( \lambda_3 \). The phases of all three components are the same as the phase of the original signal \( \tilde{x}(1) \), i.e., \( 45^\circ \).

As indicated in the first case of (3), if the signal does not reach the higher interval zone, the sub-signal in that zone is set to zero. For example, \( \tilde{x}(2) = 0.8\angle 325^\circ \) will be decomposed to \( \tilde{x}_1(2) = 0.5\angle 325^\circ, \tilde{x}_2(2) = 0.3\angle 325^\circ, \) and \( \tilde{x}_3(2) = 0 \).

The decomposed sub-signals can be represented by a matrix as

\[
\chi(n) = [\tilde{x}_1(n) \quad \tilde{x}_2(n) \quad \cdots \quad \tilde{x}_{S+1}(n)]
\]  

(4)

where

\[
\sum_{s=1}^{S+1} \tilde{x}_s(n) = \tilde{x}(n).
\]  

(5)

Since the decomposition is applied to every sample, the lengths of sub-signals are the same, and equal to that of the
original signal. For example, let us assume that the input I/Q signal in the discrete time domain has $L$ samples, and it is represented by a matrix

$$
\tilde{x} = \begin{bmatrix}
\tilde{x}(1) \\
\tilde{x}(2) \\
\vdots \\
\tilde{x}(L)
\end{bmatrix}.
$$

(6)

Using the vector decomposition approach, the above matrix can be decomposed into the following matrix:

$$
\tilde{x} = \begin{bmatrix}
\tilde{x}_1(1) & \tilde{x}_2(1) & \tilde{x}_3(1) & \cdots & \tilde{x}_{S+1}(1) \\
\tilde{x}_1(2) & \tilde{x}_2(2) & \tilde{x}_3(2) & \cdots & \tilde{x}_{S+1}(2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\tilde{x}_1(L) & \tilde{x}_2(L) & \tilde{x}_3(L) & \cdots & \tilde{x}_{S+1}(L)
\end{bmatrix}
$$

(7)

where every sample is decomposed into $S + 1$ sub-samples, represented by each row of the matrix.

B. DPD Model Selection and Output Recombination

After the input signal is decomposed into sub-signals, we now need to choose analytic functions for constructing the DPD. In order to simultaneously compensate static distortion and memory effects in RF PAs, an efficient DPD was proposed in [4], where the DPD model employed is derived from the dynamic deviation reduction-based Volterra series proposed in [5]

$$
y(n) = \sum_{p=1}^{P} h_{p,0}(0,\ldots,0)x^p(n) + \sum_{p=1}^{P} \left[ \sum_{r=1}^{P} h_{p,r}(0,\ldots,0,i_1,\ldots,i_r) \prod_{j=1}^{r} x(n-i_j) \right]
$$

where $x(n)$ and $y(n)$ are the input and output, respectively. $P$ is the order of nonlinearity and $M$ represents memory length. In this representation, the input elements are organized according to the order of dynamics involved in the model, with a variable $r$ introduced to represent the order of the dynamics, and $h_{p,r}(0,\ldots,0,i_1,\ldots,i_r)$ is the Volterra kernel with $p$th-order nonlinearity and $r$th-order dynamics. Since the effect of dynamics tends to fade with increasing order in many real PAs, the high-order dynamics can be removed by setting the value of $r$ to a small number, which leads to a significant simplification in model complexity. As shown in [4], this model can be directly adopted to represent the DPD function because DPD often only needs to compensate static nonlinearities and low-order dynamics, which are the dominant distortion induced by the PA.

In this study, we use the low-pass equivalent format of the first-order dynamic truncation of the Volterra model in (8) to construct the DPD sub-functions for each sub-signal. For instance, for the $s$th sub-signal $\tilde{x}_s(n)$ in the zone $s$, its DPD sub-function can be written as

$$
\tilde{u}_s(n) = G_s \left[ \tilde{x}_s(n) \right] = \sum_{k=0}^{P_{s-1}} \sum_{i=0}^{M_s} \tilde{g}_{s,2k+1+i}(i) |\tilde{x}_s(n)|^{2k} \tilde{x}_s(n-i)
$$

(9)

where $\tilde{x}_s(n)$ and $\tilde{u}_s(n)$ are the input and the output, respectively, and $\tilde{g}_{s,2k+1+i}(i)$ is the complex Volterra kernel of the DPD. $P_s$ is an odd number representing the order of nonlinearity and $M_s$ represents memory length. In each zone, $P_s$ and $M_s$ can be flexibly chosen according to the characteristics of the PA in that specific region. This enables the distinct behavior of the PA in the different regions to be accurately characterized, and therefore, the distortion caused by this nonlinear behavior can be effectively compensated. Note that, to reduce model complexity, no cross terms are considered in this model, namely, the output of each zone only depends on the sub-signal within that zone.

Finally, the total output of the DPD can be obtained from the sum of the outputs of all sub-functions

$$
\tilde{u}(n) = \sum_{s=1}^{S+1} \tilde{u}_s(n) = \sum_{s=1}^{S+1} \left\{ \sum_{k=0}^{P_{s-1}} \sum_{i=0}^{M_s} \tilde{g}_{s,2k+1+i}(i) |\tilde{x}_s(n)|^{2k} \tilde{x}_s(n-i) + \sum_{k=1}^{P_{s-1}} \sum_{i=1}^{M_s} \tilde{g}_{s,2k+1+i-2}(i) |\tilde{x}_s(n)|^{2k-1} \tilde{x}_s(n-i) \right\}
$$

(10)

An overview of the waveform decomposition and recombination process is outlined in Fig. 6. The original signal (on the left) is first decomposed into sub-signals located in several interval zones, corresponding to different input power levels. Each sub-signal (in the center) is processed using the dynamic deviation reduction-based Volterra function within its interval zone. All outputs of the sub-functions (on the right) are recombined together to produce the final predistorted output. We call
this DPD model the decomposed piecewise Volterra series. Although multiple sub-functions are involved in this model, the total number of parameters can be kept reasonably small since nonlinearity changes are relatively small within each interval zone, and therefore, the nonlinear order \( P_2 \) in each sub-function can be set to small values, e.g., third or fifth order.

This input-decomposition and output-recombination process is fundamentally different from the normal piecewise curve fitting. In the curve fitting, as shown in Fig. 4, the input signals are only sorted into several segments or bins, and separated in time. The value of each individual samples does not change, i.e., each sub-function still uses the original value of the signal as input. For instance, for curve fitting using the samples, \( \hat{x}(1) \) and \( \hat{x}(2) \) in subsection A, these values will be directly sent to branch 3 and branch 2 separately, and processed by sub-function \( F_3[\cdot] \) and \( F_2[\cdot] \), respectively. The final output is obtained from branch 3 for sample \( \hat{x}(1) \), i.e., \( \hat{y}[\hat{x}(1)] = F_3[\hat{x}(1)] \), and from branch 2 for sample \( \hat{x}(2) \), \( \hat{y}[\hat{x}(2)] = F_2[\hat{x}(2)] \).

However, in vector decomposition, all input samples are decomposed into smaller valued signals, and this takes place at every sampling point. Both \( \hat{x}(1) \) and \( \hat{x}(2) \) are first decomposed into the three sub-samples \( \hat{x}_1(1) \), \( \hat{x}_2(1) \), \( \hat{x}_3(1) \), \( \hat{x}_1(2) \), \( \hat{x}_2(2) \), \( \hat{x}_3(2) \), and each sub-sample is then processed by corresponding sub-functions \( G_1[\cdot] \), \( G_2[\cdot] \), \( G_3[\cdot] \). Finally, the outputs of all three branches are summed together to produce the final output, i.e., \( \hat{y}[\hat{x}(1)] = \sum_{i=1}^{3} G_i[\hat{x}_i(1)] \) and \( \hat{y}[\hat{x}(2)] = \sum_{i=1}^{3} G_i[\hat{x}_i(2)] \).

We can see that, in this case, the outputs for both \( \hat{x}(1) \) and \( \hat{x}(2) \) simultaneously involve three branches in parallel, while only one branch is involved for each input sample at a certain time sampling point in the curve-fitting case.

As mentioned earlier, there are discontinuities in the curve fitting due to possible differences in the outputs at joint points. In the vector decomposition process, the input signals are decomposed into smaller sub-signals, i.e., each sub-signal is only equal to part value of the original one (and the sum of the sub-signals restores the original signal). In this case, the origin of the vertical axis in each interval coordinate is shifted from \( \lambda_n \) to zero. In other words, the sub-signal in a certain interval zone only takes into account the interval value of the original signal that falls in that interval zone, rather than the entire value of the original signal in the curve-fitting case. This means that the minimum value of each sub-signal is zero and the output of every intervals also starts from zero [assuming there is no constant term in (9)]. The final output is obtained by recombining these sub-outputs. This guarantees that the final output is continuous at joint points.

Using piecewise curve fitting, it is very difficult to embed memory effects into the model since the sub-samples are processed in series in time and by different sub-functions. In this new DPD model, memory can be easily taken into account because the sub-signals are processed in parallel in time. As shown in (9), the samples with time delays (memory terms) can be included in each sub-functions, and then included in the final output, shown in (10). Therefore, memory effects and static nonlinearities induced by the PA can be simultaneously compensated by employing this DPD model.

More importantly, despite the fact that the input signal is decomposed into several sub-signals and more than one Volterra function is employed, the output of this model is still linear with respect to its coefficients, i.e., the final output of the DPD \( \hat{y}(n) \) is linear in relation to all coefficients \( \hat{y}_{k}[\hat{x}(n)] \), as shown in (10). This important feature implies that this nonlinear model can be extracted by using linear system identification algorithms, e.g., least squares (LS), and all parameters can be extracted by using a single estimation in one matrix. This significantly simplifies the model extraction process and also minimizes model estimation errors. In the case of curve fitting, the coefficients of each sub-function are normally identified separately, which may introduce extra errors.

C. Parameter Extraction and System Implementation

Although multiple sub-functions are involved and the core DPD model becomes different, the structure of the overall system can be kept similar to that given in [4] and the new DPD system can be characterized by using the same characterization process presented in [4].

One may notice that, during the vector decomposition, the signal bandwidth is expanded, i.e., the bandwidth of sub-signals is wider than that of the original signal due to hard shoulders, which appear on the edges of thresholds, as shown in Fig. 6. This bandwidth expansion, however, does not affect the sampling requirement for the DPD system. This is because that the signal decomposition process only takes place in the digital domain. Before reconstructing the final output, the outputs of sub-branches (predistorted sub-signals) are recombined. From the system-level perspective, the DPD is still a single-input and single-output system. To avoid aliasing effects, the sampling requirement only depends on the bandwidth of the final output of the DPD, which corresponds to the bandwidth expansion caused by the PA nonlinearity or the order of the nonlinearity to be compensated, as discussed in [4].

The overall ET system also can be considered as a “black-box” band-limited system with a single I/Q input and a single I/Q output, even though there are two paths, RF and envelope, inside the system. The under-sampling theory applied in [4] can also be applied for data acquisition and the \( p \)-th order post-inverse can be used for model extraction here. In the final DPD implementation, the original signal is over-sampled to the required sampling rate and zeros are padded into the extracted Volterra kernels to interpolate the parameters to avoid aliasing effects. The same characterization and implementation procedures presented in [4] can be used in characterizing this new system. For the sake of completeness, we briefly re-describe this process again as follows.

The complex I/Q data streams are first captured from the input and output of the PA through time-domain stimulus-response measurements at a low sampling rate. After time alignment and normalization, the input and output data are swapped, namely, the output of the PA is used as the input of the DPD, while the input of the PA is treated as the expected output of the DPD. The parameters of the DPD, i.e., Volterra kernels, are then extracted using LS estimation. To avoid aliasing effect, the original input must be over-sampled and zeros must be inserted into the extracted Volterra kernels to match the higher sampling rate, corresponding to the order of the nonlinearities to be compensated, before they are loaded into the DPD. Finally, the system...
is operated in an open-loop topology, in which the over-sampled original input signal is first predistorted by the DPD, and then modulated and up-converted to the RF frequency, and finally transmitted by the PA. The block diagram of this characterization process is shown in Fig. 7 and more details can be found in [4].

V. EXPERIMENTAL RESULTS

In order to validate the proposed DPD technique, we tested two ET PAs: one was a GaN PA from Nitronex, and the other was an Si LDMOS PA from Freescale. Both PAs were operated at 2.14 GHz and excited by a WCDMA signal with 7.56-dB PAPR.

The experimental test bench was set up as shown in Fig. 8. In this test, a WCDMA signal with a chip rate at 3.84 MHz was created at baseband as complex I/Q data in MATLAB. In the RF path, in order to eliminate the effect of I/Q imbalance in the modulator, we modulated the baseband signal in MATLAB to an IF and then up-converted it to the RF band at 2.14 GHz, and finally sent the signal to the RF PA. In the envelope path, the magnitude of I/Q data was generated in MATLAB and converted to the analog domain, and amplified by an envelope amplifier. The envelope amplifier used in this study comprises a linear stage to provide a wideband voltage source and, in parallel, a switching stage to provide an efficient current supply. The output voltage of the envelope amplifier follows the input envelope signal with help of an operational amplifier. The current is supplied to the drain of RF amplifier from both the linear stage and the switching stage through current feedback, which senses the current flowing out of the linear stages and turns the switch on/off [8].

For model extraction, the output of the RF PA is down-converted to the IF band and converted to the digital domain, and finally demodulated to the baseband in MATLAB.

As in [4], around 10000 I/Q samples were recorded with a sampling rate at 15.36 MHz, i.e., four samples per chip. After time alignment and normalization, 2000 samples were used for parameter extraction, while the remaining 8000 different samples were used for system performance evaluation in separated measurements. After the system characterization, to avoid aliasing in the final DPD system, the original input signal was over-sampled by a factor of seven, i.e., with a sampling rate at 107.52 MHz, and zeros were inserted into the previously extracted Volterra kernels to interpolate the model parameters to match the higher sampling speed. The DPD system was then operated as an open-loop topology, termed the “memory DPD” in the following text and results.

For comparison, a “memoryless DPD” was also employed in the tests. It was a polynomial-function-based model, but also using the vector decomposition approach, namely, the memory length $M_\mu$ was set to zero in (10). Another memoryless DPD model based on the polynomial curve fitting, called “curve-fitting DPD,” was also extracted and implemented in the system.
Fig. 10. AM/AM and AM/PM performance of the GaN PA without DPD and with memoryless DPD.

A. GaN PA

The first test was a high-power GaN base-station amplifier with dynamic ET. The PA was operated at 2.14 GHz, and excited by a WCDMA signal with average output power at 35.5 W. As shown earlier in Fig. 2, this PA had distinct characteristics in different power levels. In this test, the magnitude threshold was set as \(0.03, 0.15, 0.5\) for the normalized measurement data and the corresponding nonlinearity order was chosen as \(3, 3, 5, 5\). The memory length was set to two delay taps, corresponding to the duration of a half WCDMA chip. The total number of parameters was 42. Fig. 9 shows the AM/AM and AM/PM performance with and without the “memory DPD,” expressed in terms of the magnitude and phase measured for the output I/Q signal versus the corresponding instantaneous input value. It can be clearly seen that, with proposed DPD, both gain expansion and gain compression were compensated and the output magnitude linearly increased with the input. Before DPD, the phase changes were up to 25°, while after DPD, the phase changes were dropped to within 2°. The memory effects were also significantly reduced after DPD.

Fig. 11. AM/AM and AM/PM performance of the GaN PA without DPD and with curve-fitting DPD.

For comparison, the results from the “memoryless DPD” are presented in Fig. 10, where we can see that although the nonlinearities have been mostly compensated, nevertheless a large proportion of memory effects still remain. The memory effects also remain in the “curve-fitting” case, shown in Fig. 11. The spectrum plots are shown in Fig. 12, and Table I gives the adjacent channel power ratio (ACPR) and normalized root mean square errors (NRMSEs) [4] performance of the system. We can see that, with the “memory DPD,” nonlinear distortion induced by the PA was almost completely removed, with the ACPR reduced to around -58 dBc, while the NRMSE fell to only 1.96%. The “memoryless DPD” can remove most static nonlinearities, but the NRMSE can only be reduced to 4.58% at best in this case and some out-of-band distortion still remains. The “curve-fitting” DPD performed even worse, which may due to discontinuities occurring in the model.

B. LDMOS PA

The second test was an LDMOS PA with ET. The PA was also operated at 2.14 GHz, and excited by a WCDMA signal with average output power at 36.5 W. In this PA, the AM/AM distortion mainly arose in the low-power level. As shown in Fig. 13(a),
there were large gain reduction and significant memory effects when the input magnitude became low. While the input power increased, the output then almost linearly increased. However, the phase changed dramatically over the overall power range, from 10° in the lower end to 60° in the maximum power level. This caused significant distortion to the time-domain waveform, e.g., the NRMSE performance became very bad, up to 44.9%.

In this test, the threshold was set as \(0.05, 0.2, 0.3\) and the corresponding nonlinearity order was chosen as \(3, 3, 3\). The memory length was set to one delay, which equaled the duration of a quarter of a WCDMA chip. The total number of parameters was 20. Fig. 13 shows the measured AM/AM and AM/PM performance with and without the “memory DPD.” It can be seen that, despite the quite different PA behavior in this case, the distortion (both static nonlinearities and memory effects) could also be successfully compensated by employing the proposed DPD model. Similar to the previous GaN case, the “memoryless DPD” and the “curve-fitting DPD” can only compensate static nonlinearities, but memory effects still remain, although these results are not shown here. The spectrum plots are shown in Fig. 14, and ACPR and NRMSE performance are given in Table II. From these results, we can see that the “memory DPD” achieved outstanding performance. The ACPR in the first adjacent channel was reduced by over 30 dB from \(-27\) to \(-57\) dBc and NRMSE was reduced from 44.9% to 2.10%. However, both the “memoryless DPD” and the “curve-fitting DPD” only achieved limited performance.
VI. CONCLUSION

A novel DPD model for ET PAs has been presented in this paper. With the new vector threshold decomposition technique, the input complex envelope signal is first decomposed into several sub-signals and processed by dynamic deviation reduction-based Volterra series separately, and finally recombined together to produce the predistorted output. Experimental results have shown that, by employing this new piecewise type of DPD model, even when using only a small number of parameters, the distinct characteristics of the ET system in different power levels can be accurately modeled, and the distortion caused by the system nonlinear behavior can be effectively compensated.

This DPD model can be easily characterized by using discrete time-domain measurements and can be simply implemented in real digital circuits. Although we have only demonstrated its performance in the ET system in this paper, the model can be employed to linearize a wide range of PAs, especially those exhibiting distinct characteristics at different power levels.

REFERENCES

Chin Hsia received the M.Sc. (Eng.) degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1999, and is currently working toward the Ph.D. degree at the University of California at San Diego (UCSD), La Jolla.

In 2001, he joined the MBT Corporation, where he was involved in the field of optical microelectromechanical systems (MEMS). From 2002 to 2004, he was with the Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C., where he was involved with their nanosystem metrology project. Since 2004, he has been a Research Assistant with UCSD. His research interests include RF and analog circuit design for diverse fields such as microelectromechanical systems (MEMS), power supply systems, and high-efficiency transmitters for wireless communications.

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

He was subsequently involved with microwave subsystem development with Plessey Research (Caswell) U.K., prior to returning to UCD in 1980. He is currently a Professor of electronic engineering and the Chair of Electronic Engineering with UCD. His research interests are in the fields of nonlinear modeling and characterization techniques at the device, circuit and system levels. He also has interests in nonlinear simulation algorithms and several areas of microwave subsystem design and applications. He has been published widely in the international scientific literature within these fields.

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

Donald F. Kimball (S’82–M’83) was born in Cleveland, OH, in 1959. He received the B.S.E.E. degree (summa cum laude) with distinction and M.S.E.E. degree from The Ohio State University, Columbus, in 1982 and 1983, respectively.

From 1983 to 1986, he was a TEMPEST Engineer with the Data General Corporation. From 1986 to 1994, he was an Electromagnetic Compatibility Engineer/Manager with Data Products New England. From 1994 to 1999, he was a Regulatory Product Approval Engineer/Manager with Qualcomm Inc. From 1999 to 2002, he was a Research and Technology Engineer/Manager with Ericsson Inc. Since 2003, he has been a Principal Development Engineer with Cal(IT)2, University of California at San Diego, La Jolla. He holds four U.S. patents with two patents pending associated with high-power RF amplifiers (HPAs). His research interests include HPA envelope elimination and restoration (EER) techniques, switching HPAs, adaptive digital predistortion, memory-effect inversion, mobile and portable wireless device battery management, and small electric-powered radio-controlled autonomous aircraft.

Peter M. Asbeck (M’75–SM’97–F’00) received the B.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1969 and 1975, respectively.

He then joined the Sarnoff Research Center, Princeton, NJ, and Philips Laboratory, Briarcliff Manor, NY, where he was involved in the areas of quantum electronics and GaAlAs/GaAs laser physics. In 1978, he joined the Rockwell International Science Center, where he was involved in the development of high-speed devices and circuits using III-V compounds and heterojunctions. He pioneered efforts to develop heterojunction bipolar transistors based on GaAs/GaAs and InAlAs/InGaAs materials. In 1991, he joined the University of California at San Diego, where he is the Skyworks Chair Professor of the Department of Electrical and Computer Engineering. He has authored or coauthored over 300 publications. His research interests are the development of high-performance transistor technologies and their circuit applications.

Dr. Asbeck is a member of the National Academy of Engineering (NAE). He has been a Distinguished Lecturer of the IEEE Electron Device Society and the IEEE Microwave Theory and Techniques Society (IEEE MTT-S). He was the recipient of the 2003 IEEE David Sarnoff Award for his work on heterojunction bipolar transistors.