Digital Predistortion Using Extended Magnitude-Selective Affine Functions for 5G Handset Power Amplifiers With Load Mismatch

Xiaoyu Wang[ID], *Graduate Student Member, IEEE*, Yue Li[ID], *Member, IEEE*, and Anding Zhu[ID], *Fellow, IEEE*

Abstract—Load mismatch often occurs in radio frequency (RF) power amplifiers (PAs) in handset, which can complicate the nonlinear behavior of the transmitter, particularly when the voltage standing wave ratio (VSWR) is high. To effectively linearize handset PAs with load mismatch, in this article, we propose a low-complexity digital predistortion (DPD) model. After carefully analyzing the characteristics of handset transmitters with standing waves, we extend the magnitude-selective affine (MSA) function to construct nonlinear modeling terms in the DPD model. The new terms can effectively compensate for the distortion caused by load mismatch. Experimental results demonstrate that the proposed extended MSA (EMSA) model not only improves the modeling performance but also lowers the hardware complexity considerably, compared with the conventional models. The new model can thus serve as a good candidate for DPD deployment in 5G handset transmitters, where nonlinear behavior is more complicated and hardware resource is highly constrained.

Index Terms—Digital predistortion (DPD), 5G, handset transmitter, load mismatch, power amplifier (PA), standing waves.

I. INTRODUCTION

POWER amplifier is one of the most important components in radio frequency (RF) systems. It consumes a large portion of energy and dominates the nonlinearity of wireless transmitters. The complex nonlinear behavior of power amplifiers (PAs) in the transmitters can lead to significant distortions and deteriorate the signal quality [1], [2]. To maintain linearity, digital predistortion (DPD) has been widely adopted as an effective linearization technique in modern communication systems, especially in high-power base stations [3].

Over the past decades, many DPD models have been proposed [4], [5]. The majority of models are developed by pruning the Volterra series, including memory polynomials (MP) [6], generalized MP (GMP) [7], and dynamic deviation reduction (DDR) [8]. The pruned Volterra models

have intrinsic difficulties in dealing with strong nonlinearities, because they need high-order polynomial functions with a large number of coefficients, which can lead to high hardware complexity. To alleviate this issue, lookup table (LUT) [9], [10] and spline-based [11] methods have been adopted to reduce the implementation complexity. Other behavioral model expressions have also been proposed. In particular, canonical piecewise linear (CPWL) function-based models [12], including decomposed vector rotation (DVR) model [13], [14] and magnitude-selective affine (MSA) function-based models [15], [16], have shown a great potential to achieve good accuracy with low hardware cost. Recently, piecewise models have also been developed to solve the tradeoff between linearization performance and complexity [17]–[19].

In typical handsets, the transmit power is much lower than that of base stations, which leaves little power budget for DPD. The linearity requirement is also less stringent [20]. DPD thus has not been widely applied to the existing handset transmitters. However, with continuous demands for wide bandwidths and high data rates, linearity becomes a serious issue in 5G handset and thus DPD may play an important role in the design of future handset transmitters. Yet, research on DPD techniques in handset transmitters is still in an early stage. The existing solutions mainly focus on reducing hardware complexity [21] and largely concentrate on LUT-based models with least mean squares (LMS)-like adaptation algorithms [22], [23]. While the complexity of these solutions is low, they only offer limited linearization performance, which can hardly satisfy the linearity requirement in wideband systems.

Another prominent problem for DPD in handset transmitters is the tolerance for load mismatch or high voltage standing wave ratio (VSWR) [24]. Various factors influence the load of PA output, e.g., the blockage of antennas by the position of hands. The resulting load variations lead to suboptimal impedance at PA output, which will degrade the performance of PAs. This problem becomes increasingly severe, due to the wider signal bandwidth in 5G. Whereas different VSWR protection techniques have been considered to fight against the load mismatch during the PA design stage [25]–[30], they may be insufficient to fully overcome the performance degradation caused by load mismatch. Antenna tuners can be applied in handsets to compensate the variations of VSWR. However, the

existing antenna tuners generally have limited tuning range and resolution, and impedance change caused by external environment, e.g., holding position of handset, is difficult to detect [31]. The tuning speed can also be a bottleneck, as it may fail to track the fast variation of the antenna impedance. Moreover, when MIMO arrays are deployed, the integration of bulky antenna tuners becomes prohibitive [32]. Considering all these practical difficulties, it is very likely that the PA needs to cope with a relatively high VSWR in certain cases.

To model the PA behavior under load mismatch conditions, the authors of [33]–[35] proposed two-port behavioral models which take the reflective wave into consideration. In [36], a feedback finite impulse response (FIR) filter cross-term MP model was proposed to compensate for PA nonlinearity with load mismatch. In [37], the authors compared the performance of three widely used DPD models when dealing with different mismatch conditions, which suggested that models without cross-terms can hardly realize effective linearization, especially for load-modulated PAs. Because the PAs exhibit very different characteristics under load mismatch, it is important to carry out new investigations and design new models specifically for handset PAs with load mismatch.

In this article, we propose a low-complexity DPD method for handset transmitters considering the effect of load mismatch. Based on a thorough analysis of handset PAs with output standing waves, we find that additional cross-terms are needed to accurately describe its behavior under load mismatch conditions. To fit more cross-terms in the new DPD model and maintain low implementation complexity, we transform the complicated nonlinear terms into hardware-friendly blocks by extending the MSA functions. With the extension, the zone selection in the conventional MSA function can work with any real-valued quantity. Hardware complexity can be further reduced by systematically examining the design and hardware issues of the proposed DPD model. We demonstrate, through experimental results, that the proposed DPD method can effectively compensate for the complicated nonlinearities and, in the meantime, maintain low implementation complexity.

The rest of this article is organized as follows. In Section II, the behavioral modeling of the PA with load mismatch is analyzed. Section III introduces the DPD model design and the implementation. Subsequently, Section IV reports the experimental results, followed by a conclusion in Section V.

II. BEHAVIORAL MODELING OF PA WITH LOAD MISMATCH

In wireless base stations, an isolator, or an equivalent component, is usually used to separate PA and antenna to ensure that the optimum impedance condition is provided to the PA, which helps to maintain good linearity and efficiency. However, in handsets, the isolator is usually eliminated due to cost and space limit. The handset PAs thus may suffer from significant load mismatch, leading to additional nonlinear distortions.

The incident and reflected signals of a general PA block are shown in Fig. 1, where a_1 and b_1 are the incident and
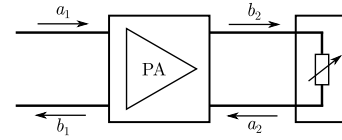

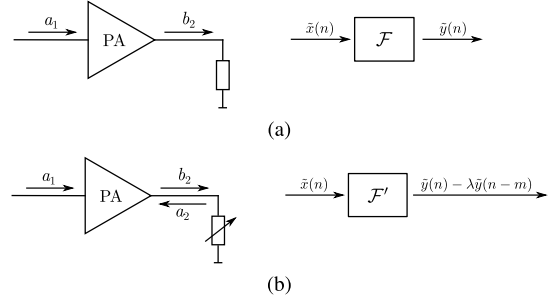Fig. 1. Incident and reflected signals of PA.

Fig. 2. Output of PA and the corresponding PA model. (a) Without standing waves. (b) With standing waves.

reflected signals of PA input, respectively, while a_2 and b_2 are the incident and reflected signals of PA output, respectively. As displayed in Fig. 2(a), when PA does not have the problem of load mismatch, the incident signal at the output port can be ignored, i.e., $a_2 = 0$, and only b_2 is transmitted. Therefore, the baseband equivalent of the transmitted signal can be simply represented as

$$\tilde{y}(n) = \mathcal{F}\big(\tilde{x}(n)\big) \tag{1}$$

where $\mathcal{F}$ is the PA model without mismatch, and $\tilde{x}(n)$ and $\tilde{y}(n)$ are the baseband equivalent of PA input and transmitted output signals. To linearize the system, an inverse nonlinear function $\mathcal{G}$, presented in Fig. 3(a), can be found by building a postinverse model

$$\tilde{x}(n) \approx \mathcal{G}\big(\tilde{y}(n)\big). \tag{2}$$

The postinverse function $\mathcal{G}$ can be directly used in DPD by swapping input and output.

However, if the output impedance is not well matched, part of the PA output signal will be reflected back by the load, as shown in Fig. 2(b), leading to degradation of efficiency and linearity. In this case, we have

$$a_2(t) = \Gamma_0 b_2(t - \tau) \tag{3}$$

where Γ_0 is the reflection coefficients and τ represents a time delay. Because of the complex PA behavior and load conditions, the signal a_2 may appear as part of the transmitted signal. The final radiated signal depends on both a_2 and b_2. The baseband equivalent of the radiated signal can then be expressed as

$$\tilde{y}'(n) = \tilde{y}(n) - \lambda \tilde{y}(n - m) \tag{4}$$

where λ is a coefficient and m shows the equivalent time delay in the baseband. This indicates that, besides the original output signal, a delayed copy is also present. In addition, the composite signal will directly affect the PA nonlinearity.
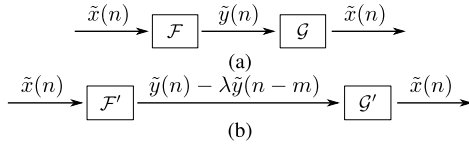
Fig. 3. PA and postinverse model. (a) Without standing waves. (b) With standing waves.

Fig. 4. Representation of MSA function. (a) Conventional view. (b) Dual-input view.

To linearize $\tilde{y}'(n)$, as displayed in Fig. 3(b), a postinverse model $\mathcal{G}'$ can be expressed as

$$\tilde{x}(n) \approx \mathcal{G}'\big(\tilde{y}'(n)\big) = \mathcal{G}'\big(\tilde{y}(n) - \lambda\tilde{y}(n - m)\big). \quad (5)$$

Comparing (2) and (5), we can see that, under load mismatch, to build the inverse function, a composite signal is required in the input. For instance, for the third-order polynomials, instead of $|\tilde{x}(n)|^2\tilde{x}(n)$, we should use a new term $|\tilde{x}(n) + \lambda\tilde{x}(n - m)|^2\big(\tilde{x}(n) + \lambda\tilde{x}(n - m)\big)$. Since

$$|\tilde{x}(n) + \lambda\tilde{x}(n - m)|^2 = |\tilde{x}(n)|^2 + |\lambda|^2|\tilde{x}(n - m)|^2$$
$$+ 2\text{Re}[\lambda^*\tilde{x}(n)\tilde{x}^*(n - m)] \quad (6)$$

we have

$$|\tilde{x}(n) + \lambda\tilde{x}(n - m)|^2\big(\tilde{x}(n) + \lambda\tilde{x}(n - m)\big)$$
$$= |\tilde{x}(n)|^2\tilde{x}(n) + |\lambda|^2|\tilde{x}(n - m)|^2\tilde{x}(n)$$
$$+ \lambda|\tilde{x}(n)|^2\tilde{x}(n - m) + \lambda|\lambda|^2|\tilde{x}(n - m)|^2\tilde{x}(n - m)$$
$$+ 2\text{Re}[\lambda^*\tilde{x}(n)\tilde{x}^*(n - m)]\tilde{x}(n)$$
$$+ 2\lambda\text{Re}[\lambda^*\tilde{x}(n)\tilde{x}^*(n - m)]\tilde{x}(n - m). \quad (7)$$

As a result, $|\tilde{x}(n) + \lambda\tilde{x}(n - m)|^2\big(\tilde{x}(n) + \lambda\tilde{x}(n - m)\big)$ can be transformed into the following terms:

$$\begin{cases} |\tilde{x}(n)|^2\tilde{x}(n) \\ |\tilde{x}(n - m)|^2\tilde{x}(n) \\ |\tilde{x}(n)|^2\tilde{x}(n - m) \\ |\tilde{x}(n - m)|^2\tilde{x}(n - m) \\ \text{Re}[\tilde{x}(n)\tilde{x}^*(n - m)]\tilde{x}(n) \\ \text{Re}[\tilde{x}(n)\tilde{x}^*(n - m)]\tilde{x}(n - m). \end{cases} \quad (8)$$

The resulting terms show that, to linearize a PA with load mismatch, not only the conventional polynomial terms of $\tilde{x}(n)$ but also the new terms of $\text{Re}[\tilde{x}(n)\tilde{x}^*(n - m)]$ must be included.

For simplicity, only one delayed signal is considered in the derivation in (4) to show how the reflected signals are involved in the model. In the practical scenario, the reflection coefficient of antenna can be frequency dependent, which means that more than one delayed signal shall be included in (4), and these terms may be represented as the weighted sum of the reflected signals with different delays. The derivation can be conducted in the same way as that from (5) to (7). The other way to model memory effects caused by the frequency-dependent response of antenna is to include multiple delayed signals in the model. To be more specific, terms in (8) with different m values can be included in the model so that the final model is expected to model the memory effect caused by the antenna properly.
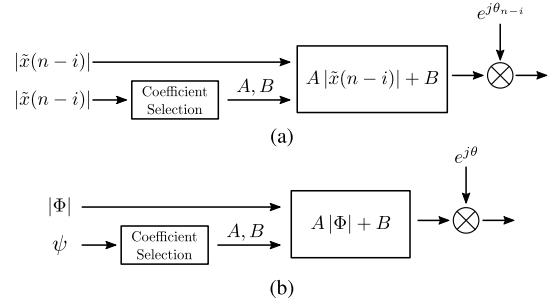
III. DPD Model Design and Implementation

From the analysis presented in Section II, we can see that the nonlinearity of the PA with load mismatch is more complicated than that of the usual case and requires extra nonlinear terms in models. Moreover, to apply DPD on handset transmitters, low complexity is a key consideration, since there is limited budget for DPD in handset due to power and cost constraints. Directly implementing (8) as that of the Volterra series models is thus not feasible in this situation, because high-order polynomials lead to high complexity. In this work, to realize the linearization of the handset transmitters with load mismatch, we introduce a low-complexity DPD model by extending the MSA function to fulfill the requirements.

A. Review of MSA Function

The MSA functions [15] build nonlinearity in a piecewise manner for an efficient operation. The piecewise linear operation is achieved by selecting different coefficients for the affine function $\tilde{x}(n - i)$, according to the input magnitude $|\tilde{x}(n - i)|$. The MSA function can be written as

$$\tilde{u}(n - m)$$
$$= \begin{cases} (A_1|\tilde{x}(n - m)| + B_1)e^{j\theta_{n-m}}, & |\tilde{x}(n - m)| < \beta_1 \\ (A_2|\tilde{x}(n - m)| + B_2)e^{j\theta_{n-m}}, & \beta_1 \le |\tilde{x}(n - m)| < \beta_2 \\ \quad\quad \vdots & \quad\quad \vdots \\ (A_K|\tilde{x}(n - m)| + B_K)e^{j\theta_{n-m}}, & |\tilde{x}(n - m)| \ge \beta_{K-1}. \end{cases}$$
$$(9)$$

As shown in Fig. 4(a), both the affine function and the coefficient selection process employ $|\tilde{x}(n - i)|$ as input. In fact, the MSA function can be understood as a dual-input nonlinear operator, because the selection of model coefficients can be independent of the affine function expression [16]. For example, if we change the input of coefficient selection to $|\tilde{x}(n)|$, we can obtain a different model term

$$\tilde{u}(n - m)$$
$$= \begin{cases} (A_1|\tilde{x}(n - m)| + B_1)e^{j\theta_{n-m}}, & |\tilde{x}(n)| < \beta_1 \\ (A_2|\tilde{x}(n - m)| + B_2)e^{j\theta_{n-m}}, & \beta_1 \le |\tilde{x}(n)| < \beta_2 \\ \quad\quad \vdots & \quad\quad \vdots \\ (A_K|\tilde{x}(n - m)| + B_K)e^{j\theta_{n-m}}, & |\tilde{x}(n)| \ge \beta_{K-1}. \end{cases}$$
$$(10)$$

Since the nonlinear operator simultaneously depends on two different samples, it can be directly used to build the cross-terms. Implementing the cross-terms this way is much more efficient than the polynomial models that need dedicated multipliers to mix different memory samples. To generalize this idea, as depicted in Fig. 4(b), we can express the function as

$$\mathcal{H}_K(\Phi, \psi) = \begin{cases} (A_1|\Phi| + B_1)e^{j\theta}, & \psi < \beta_1 \\ (A_2|\Phi| + B_2)e^{j\theta}, & \beta_1 \le \psi < \beta_2 \\ \quad\vdots & \quad\vdots \\ (A_K|\Phi| + B_K)e^{j\theta}, & \psi \ge \beta_{K-1} \end{cases} \tag{11}$$

where Φ is a complex-valued input signal, θ is the phase of Φ, and ψ is a real-valued splitting feature. By using different formulations for Φ and ψ, different model basis functions can be constructed. For instance, (10) can be obtained by setting

$$\begin{cases} \Phi = \tilde{x}(n - m) \\ \psi = |\tilde{x}(n)|. \end{cases} \tag{12}$$

In this new format, the input signal Φ constructs the linear part of model, including both amplitude and phase information. The nonlinearity of the model is generated from the coefficient selection process, since its gain A and bias B are varied based on the values of ψ. Thus, the MSA function in (11) can provide a similar modeling capability as polynomial terms in the form of $\psi^p\Phi$. Because the MSA function only needs to implement a simple affine function, the hardware complexity is greatly reduced.

B. Proposed DPD Model

To accurately model the nonlinear distortions caused by load mismatch while keeping low complexity, we transform the model terms in (8) with MSA functions. We obtain

$$\tilde{u}_1(n) = \sum_{m=1}^{M} \mathcal{H}_K(\tilde{x}(n), |\tilde{x}(n - m)|) \tag{13}$$

$$\tilde{u}_2(n) = \sum_{m=1}^{M} \mathcal{H}_K(\tilde{x}(n - m), |\tilde{x}(n)|) \tag{14}$$

$$\tilde{u}_3(n) = \sum_{m=1}^{M} \mathcal{H}_K(\tilde{x}(n), \text{Re}[\tilde{x}(n)\tilde{x}^*(n - m)]) \tag{15}$$

$$\tilde{u}_4(n) = \sum_{m=1}^{M} \mathcal{H}_K(\tilde{x}(n - m), \text{Re}[\tilde{x}(n)\tilde{x}^*(n - m)]) \tag{16}$$

where K and M are the number of segments and the memory depth, respectively. It is worth mentioning that, when processing $\text{Re}[\tilde{x}(n)\tilde{x}^*(n - m)]$ in (15) and (16) for coefficient selection, the threshold selection is different from those used for $|\tilde{x}(n-m)|$ in (13) and $|\tilde{x}(n)|$ in (14). $\tilde{x}(n-m)$ and $\tilde{x}(n)$ are the first-order terms, and taking the magnitude always leads to a positive number. In this case, a uniform spacing $\beta_k = k/K$ is usually used, where K is the number of segments. For example, if $K = 10$, the thresholds will be 0.1, 0.2, 0.3, etc. In contrast, $\tilde{x}(n)\tilde{x}^*(n - m)$ is a second-order term, and the value of $\text{Re}[\tilde{x}(n)\tilde{x}^*(n-m)]$ can be either positive or negative,

TABLE I
FORMULATION OF NONLINEAR TERMS USING MSA FUNCTION IN THE PROPOSED EMSA MODEL

Basis Function Type	Choice of Φ	Choice of ψ
1	$\tilde{x}(n)$	$|\tilde{x}(n - m)|$
2	$\tilde{x}(n - m)$	$|\tilde{x}(n)|$
3	$\tilde{x}(n)$	$Re[\tilde{x}(n)\tilde{x}^*(n - m)]$
4	$\tilde{x}(n - m)$	$Re[\tilde{x}(n)\tilde{x}^*(n - m)]$
5	$\tilde{x}(n - m - l)$	$|\tilde{x}(n - m)|$

because only the real part is taken. In this consideration, spacing with a quadratic curve may be used. For instance, in our test, we use the following thresholds:

$$\beta'_k = \beta_k^2 - \beta_1 \tag{17}$$

where β'_k is a quadratic function of β_k, while β_k is uniformly spaced and β_1 is the first threshold. When $K = 10$, the thresholds will be -0.09, -0.06, -0.01, etc.

In wideband systems, the nonlinearity of the PA becomes more complex. To enhance the linearization performance under wideband scenarios, more cross-terms can be added to the model. For instance, the GMP-like nonlinear terms

$$|\tilde{x}(n - m)|^p \tilde{x}(n - m - l) \tag{18}$$

can be added. These terms can be implemented by using the MSA function as

$$\tilde{u}_5(n) = \sum_{l=0}^{L} \sum_{m=0}^{M'} \mathcal{H}_K(\tilde{x}(n - m - l), |\tilde{x}(n - m)|) \tag{19}$$

where L is the cross-term delay order and M' is the memory depth for the additional terms, which may take a different value than M.

The corresponding transformations for all the terms are detailed in Table I. By aggregating all the terms, the complete model can be expressed as

$$\tilde{u}(n) = \sum_{k=1}^{5} \tilde{u}_k(n). \tag{20}$$

We refer to this model as the "extended MSA (EMSA)" model.

In practical situations, depending on the PA characteristics and the range of load variation, specific model types can be excluded from the implemented to save hardware resources. Moreover, in our prototype, all the model terms use the same number of segments. The memory depth for the first four terms is M, while the last type uses M'. The choice of setting two memory depth parameters gives more flexibility to the model and does not require too much tuning. In practical deployment, the configuration of different terms, including both memory depth and the number of segments, can be tuned individually to achieve the optimum tradeoff between performance and complexity.

As the proposed model is linear in parameters, the conventional linear system identification algorithms, such as least
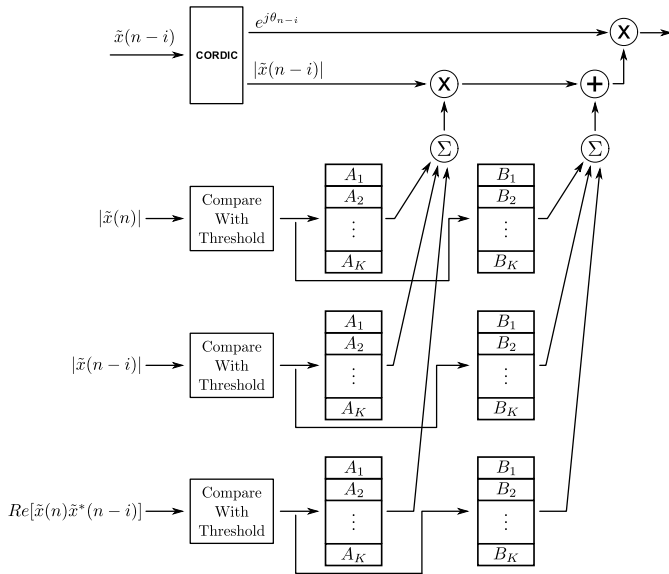
Fig. 5. Example of hardware implementation for the EMSA model.

Fig. 6. Experimental test bench.

squares (LS), can be conveniently adopted to extract the model coefficients. The model can be expressed into matrix form as

$$\mathbf{u} = \mathbf{X}\mathbf{c} \tag{21}$$

where $\mathbf{u}$ represents the DPD output, $\mathbf{X}$ includes all model basis functions, and $\mathbf{c}$ is the model coefficients.

The model can be identified by using LS as

$$\hat{\mathbf{c}} = \left(\mathbf{X}^H \mathbf{X}\right)^{-1} \mathbf{X}^H \mathbf{u}. \tag{22}$$

C. Hardware Implementation

To implement the proposed model in hardware, the input signal is first decomposed into magnitude and phase. Depending on the construction of individual terms, the correct coefficients are selected for different MSA functions. After performing the calculation of the affine function, the phase information is restored. The final output is obtained by summing up the output of all MSA blocks. An example hardware implementation is depicted in Fig. 5.

One thing to note is that, while we may need to calculate extra quantities for coefficient selection in some terms, i.e., $\mathrm{Re}[\tilde{x}(n)\tilde{x}^*(n-m)]$, there is little hardware complexity associated with it. It is because the value is only used to determine which coefficient to use, so it can be computed with very low resolution. Moreover, conceptually in the model formulation, we need to subtract all potential threshold values to locate the correct coefficients, but in actual implementation, we may not need dedicated adders or comparators. Because of the low resolution, they can be easily represented by logic functions and simplified by hardware synthesis tools.

Another important hardware optimization technique is to share the hardware resources between different MSA blocks. Since several affine functions share the same model structure, the separate coefficients A and B can be added together before multipliers. The computational complexity is further reduced
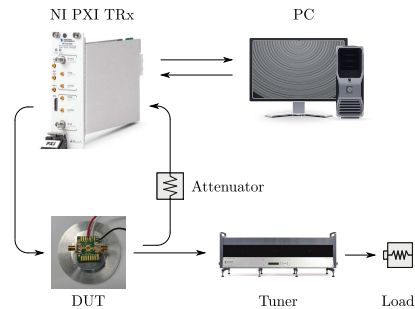
in this way. For example, consider the second term which has $\Phi = \tilde{x}(n-m)$ and $\psi = |\tilde{x}(n)|$, the fourth term which has $\Phi = \tilde{x}(n-m)$ and $\psi = \mathrm{Re}[\tilde{x}(n)\tilde{x}^*(n-m)]$, and part of the fifth term ($l = 0$) which has $\Phi = \tilde{x}(n-m)$ and $\psi = |\tilde{x}(n-m)|$. The MSA functions for different memory taps use the same affine function $(A_k|\tilde{x}(n-m)|+B_k)e^{j\theta_{n-m}}$ but have different coefficients. Thus, as shown in Fig. 5, coefficients A and B for different memory taps can be added together before multiplying with the amplitude $|\tilde{x}(n-m)|$ and the phase information $e^{j\theta_{n-m}}$. Therefore, while the complete model includes several different types of cross-terms, most of them can be combined at an early stage, and the actual hardware utilization can be kept at a low level.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

To validate the model performance, a test platform was set up as shown in Fig. 6, including PC, vector signal transceiver, PA, tuner, and attenuator. Two gallium arsenide (GaAs) class-AB PAs were used for test. PA 1 was operated at band N77 (the center frequency is 3.5 GHz) with an average output power of 22 dBm, while PA 2 was operated at band N79 (the center frequency is 4.6 GHz) with an average output power of 21 dBm. The excitation input signal was 100-MHz 16-QAM 5G new radio cyclic prefix (CP) orthogonal frequency-division multiplexing (OFDM) signal with 8-dB peak-to-average power ratio (PAPR). The system used a time division duplex (TDD) mode with 50% duty cycle. The sampling rate was 491.52 MHz. The signal was generated in MATLAB, then downloaded to the vector signal transceiver (NI PXI 5840) and finally sent to PA. A tuner (MPT 1807) was connected to PA output to control the strength of standing waves. The tuner was set to achieve VSWR = 1, VSWR = 2, and VSWR = 3. The phase of the reflection coefficients was set to 0. VSWR = 1 means that the load is matched at the output port of PA, while standing waves are present at the output port of PA when VSWR = 2 and VSWR = 3. The PA output was obtained by the coupler and fed into the vector signal transceiver. Recorded I/Q input and output samples were time-aligned and normalized before training the DPD model.

The DPD model coefficients were extracted with indirect learning architecture (ILA) in MATLAB. During the test, 50 000 samples were used for model extraction and another
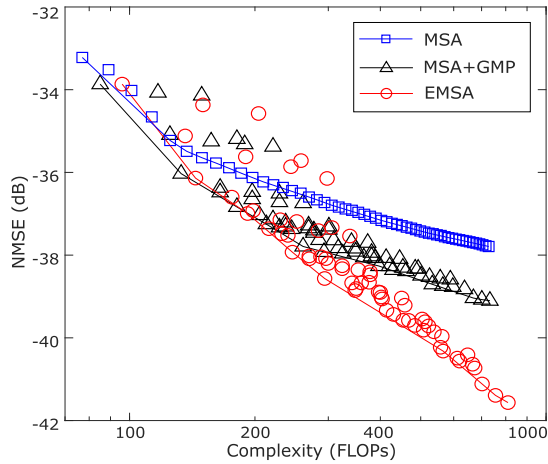
Fig. 7. NMSE performance of the proposed model using different nonlinear terms in the case of VSWR = 3.

Fig. 8. NMSE performance of GMP, MSA, and the proposed models with sweeping parameters in the case of VSWR = 3.

440 000 samples were used for performance verification. In the DPD tests, we employ both normalized mean squared error (NMSE) and adjacent channel power ratio (ACPR) as performance metrics. In the complexity analysis, we follow the same methodology as in [38] using the number of floating-point operations per sample (FLOPs).

GMP and MSA models are two popular and commonly used DPD models in the area of Volterra-based model and piecewise model, respectively. Therefore, they were tested in the experiment as baselines.

B. Forward Modeling

To explore the performance of DPD models, GMP model, MSA model, and the proposed EMSA model were first applied to forward modeling. PA 1 was operated at band N77 with 22-dBm output power and VSWR = 3.

To confirm the effectiveness of different terms in the EMSA model, we performed a parametric sweep of model parameters with different model configurations and evaluated the best performance/complexity tradeoff for each model. Before applying the sweep, all the parameters, such as memory depth, nonlinear order, length of cross-terms, etc., were increased until there was no further improvement in the performance. The sweeping range for each parameter was determined by using this criterion. The simplest model is the conventional MSA model. The second model under comparison is the EMSA model without the proposed terms (type 3 and type 4). Only type 1, type 2, and type 5 are used, and we refer this model as "MSA + GMP" which includes the GMP-like cross-terms. The last model is the full EMSA model. In the sweep, the number of segments K was set to 16. The swept parameters M and M' changed from 2 to 8, and L varied from 4 to 16. For the MSA model, the memory depth M was swept from 5 to 30 to match the complexity range.

From the modeling results shown in Fig. 7, we see that the "MSA + GMP" model outperforms the standard MSA model but performs worse than the full EMSA model. In this work, type 3/4 terms compensate for load mismatch and type 5 terms are used for wideband modeling. Thus, it demonstrates
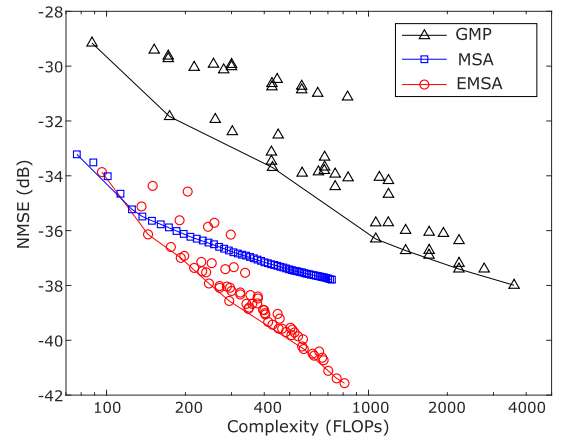
TABLE II
PARAMETERS OF NONLINEAR TERMS IN THE EMSA MODEL

No.	Choice of Φ	Choice of ψ	Config. 1	Config. 2
1	$\tilde{x}(n)$	$|\tilde{x}(n-m)|$	M=2	M=4
2	$\tilde{x}(n-m)$	$|\tilde{x}(n)|$	M=2	M=4
3	$\tilde{x}(n)$	$Re[\tilde{x}(n)\tilde{x}^*(n-m)]$	M=2	M=4
4	$\tilde{x}(n-m)$	$Re[\tilde{x}(n)\tilde{x}^*(n-m)]$	M=2	M=4
5	$\tilde{x}(n-m-l)$	$|\tilde{x}(n-m)|$	M'=2, L=4	M'=2, L=16

that both the load mismatch and wideband challenge must be addressed in order to obtain a good performance, and the proposed EMSA model is a better choice to achieve this goal.

Next, we performed the sweep over different models to compare the performance and complexity of the proposed method and the conventional models. The swept parameters of GMP model included memory depth M, polynomial order P, leading cross-term length L_1, and lagging cross-term length L_2. Specifically, we swept P from 2 to 6, changed M from 4 to 16, and varied L_1 and L_2 from 0 to 2. The sweeping range for the MSA and EMSA models was kept the same as the previous setup.

We plot the training performance of all model settings for GMP, MSA, and EMSA models in Fig. 8, where each marker represents the performance and complexity for one model configuration. The solid lines in the plot represent the best performance/complexity tradeoff for each model. The comparison shows that using GMP or MSA model alone does not show good performance and the proposed DPD model significantly outperforms the GMP and MSA models in terms of both modeling accuracy and computational complexity. By employing the proposed method, the achievable NMSE performance is further improved by nearly 4 dB within the sweeping range. The results show that including the new derived terms, i.e., the third- and fourth-type nonlinear terms in the EMSA model, significantly improves the modeling performance. Moreover, we can see that the performance improvement of GMP and MSA models is limited with increasing complexity. The best performance with GMP and

TABLE III
EXPERIMENTAL RESULTS ON PA 1

| Model	Config.	FLOP	VSWR = 1		VSWR = 2		VSWR = 3	
			NMSE(dB)	ACPR(dBc)	NMSE(dB)	ACPR(dBc)	NMSE(dB)	ACPR(dBc)
w/o DPD	N/A	N/A	-26.22	-36.95/-34.89	-21.85	-38.36/-35.88	-18.74	-39.16/-35.10
GMP	1	302	-35.30	-44.06/-44.65	-33.79	-42.62/-43.56	-32.19	-41.81/-41.68
	2	1066	-36.68	-45.19/-45.01	-36.00	-44.79/-44.54	-35.39	-44.46/-43.72
MSA	1	65	-34.00	-44.79/-44.87	-33.47	-43.80/-44.09	-32.95	-43.35/-42.80
	2	329	-37.40	-44.79/-44.88	-36.02	-43.80/-44.07	-35.01	-43.35/-42.78
Proposed	1	96	-34.30	-45.28/-45.14	-33.87	-44.60/-44.56	-33.61	-44.38/-43.78
	2	294	-37.28	-45.46/-45.36	-36.74	-45.13/-44.78	-36.38	-45.00/-44.18
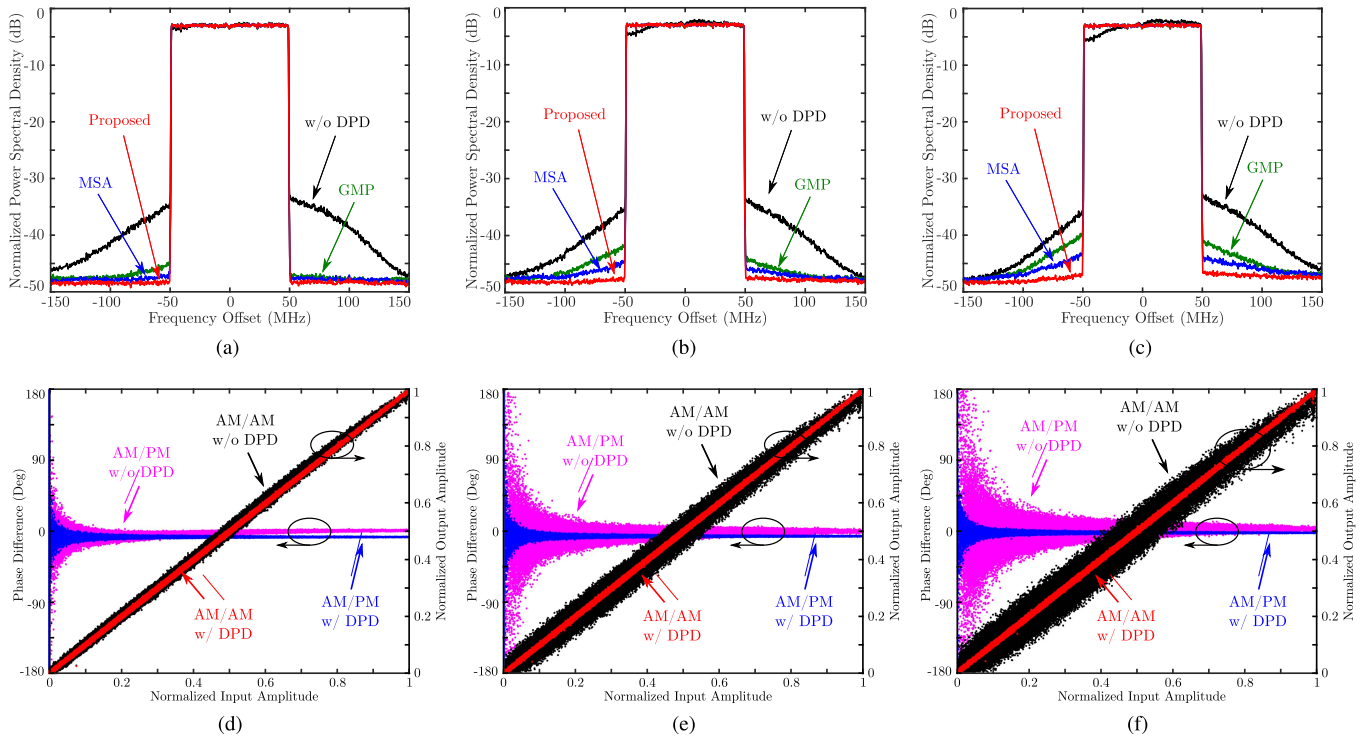
Fig. 9. Experimental results of GMP model (302 FLOPs), MSA model (329 FLOPs), and the proposed EMSA model (294 FLOPs) on PA 1, spectral comparison with (a) VSWR = 1, (b) VSWR = 2, and (c) VSWR = 3, and AM–AM and AM–PM characteristics with and without the proposed EMSA model (294 FLOPs) with (d) VSWR = 1, (e) VSWR = 2, and (f) VSWR = 3.

MSA models can be realized by the proposed method with much lower complexity. It suggests that the GMP and MSA models fail to give an efficient representation of the PA behavior under load mismatch scenarios, while the proposed model can accurately characterize PA's behavior in such a case.

C. DPD Tests on PA 1

The proposed DPD model was measured and compared with the GMP model and the MSA model on PA 1. In this comparison, all the models were tested with two different model configurations according to the solid lines in Fig. 8, so the selected configurations can reflect the best complexity-performance tradeoff for each model. To fairly compare the performance of different models, configurations of each model with similar complexity were selected. For the proposed EMSA model, the

number of segments is set to $K = 16$ in both configurations. For other parameters, configuration 1 has $M = M' = 2$ and $L = 4$, while configuration 2 has $M = 4$, $M' = 2$, and $L = 16$. The details of the model configurations are shown in Table II. For the GMP model, the first configuration of the GMP model is nonlinear order $P = 4$, memory depth $M = 8$, and no cross-terms used. This configuration of the GMP model makes it equivalent to the MP model. The second configuration of the GMP model is $P = 4$, $M = 2$, leading cross-term length $L_1 = 2$, and lagging cross-term length $L_2 = 0$. The memory length is set to be $M = 5$ in the MSA model's first configuration and $M = 27$ in the second configuration.

The spectral results of three models in the three cases of VSWR = 1, VSWR = 2, and VSWR = 3 are shown in Fig. 9(a)–(c). Fig. 9(d)–(f) displays the AM–AM and AM–PM results of the proposed EMSA model with varying

TABLE IV
EXPERIMENTAL RESULTS ON PA 2

| Model	Config.	FLOP	VSWR = 1		VSWR = 2		VSWR = 3	
			NMSE(dB)	ACPR(dBc)	NMSE(dB)	ACPR(dBc)	NMSE(dB)	ACPR(dBc)
w/o DPD	N/A	N/A	-26.53	-36.27/-36.85	-20.31	-34.30/-42.18	-16.57	-32.91/-43.48
GMP	1	302	-34.84	-44.08/-43.99	-33.68	-42.71/-43.85	-32.18	-40.64/-43.62
	2	1066	-35.57	-44.33/-44.45	-35.56	-44.07/-44.49	-35.19	-43.45/-44.57
MSA	1	65	-33.84	-44.01/-43.93	-33.71	-43.63/-43.96	-33.27	-42.48/-43.61
	2	329	-36.03	-43.99/-43.92	-35.89	-43.58/-43.99	-35.13	-42.52/-43.60
Proposed	1	96	-34.15	-44.29/-44.26	-34.13	-44.15/-44.53	-33.95	-43.63/-44.48
	2	294	-35.99	-44.35/-44.44	-36.00	-44.24/-44.55	-35.92	-43.89/-44.60
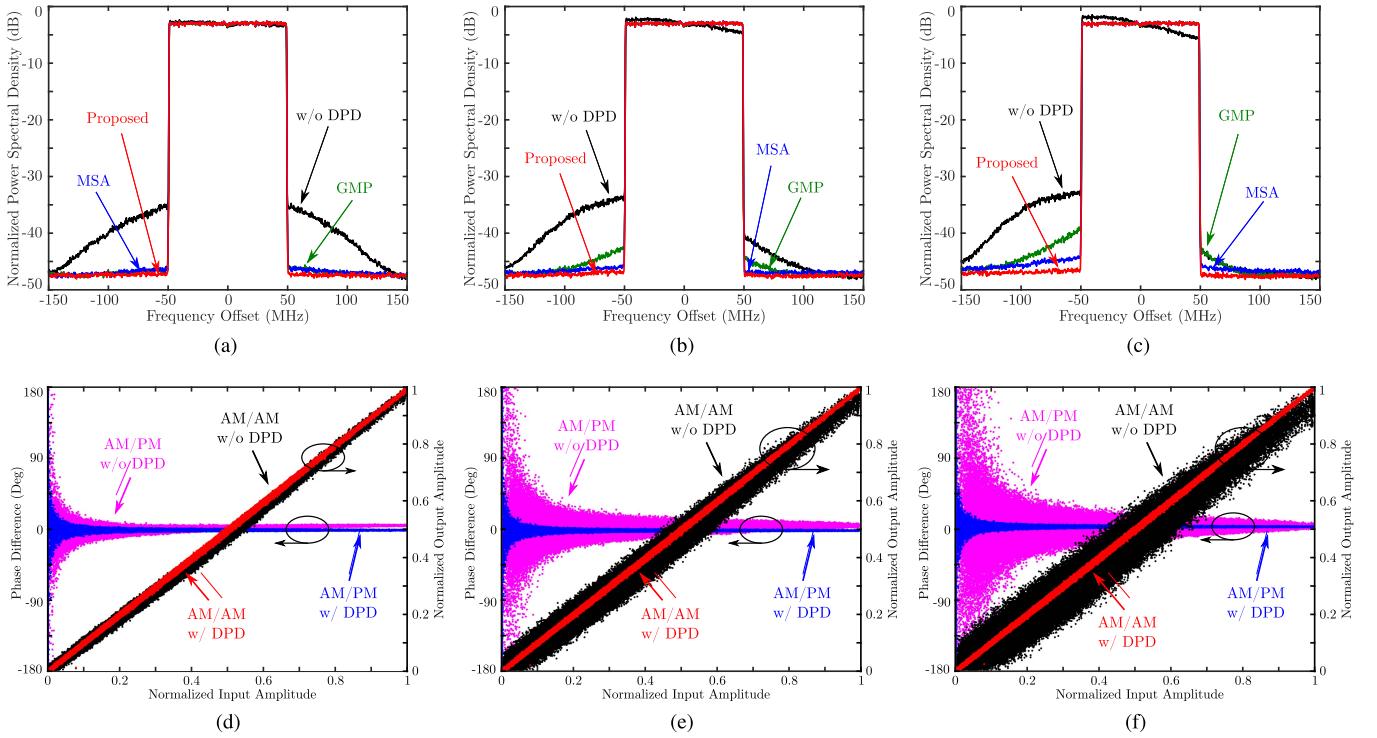
Fig. 10. Experimental results of GMP model (302 FLOPs), MSA model (329 FLOPs), and the proposed EMSA model (294 FLOPs) on PA 2, spectral comparison with (a) VSWR = 1, (b) VSWR = 2, and (c) VSWR = 3, and AM–AM and AM–PM characteristics with and without the proposed EMSA model (294 FLOPs) with (d) VSWR = 1, (e) VSWR = 2, and (f) VSWR = 3.

VSWR values. For a fair performance comparison, all the three models in the figures have similar complexity and use around 300 FLOPs. The GMP model uses 302 FLOPs, the MSA model employs 329 FLOPs, and the proposed model adopts 294 FLOPs. Seen from the spectrum and AM–AM/AM–PM plots in the case with standing waves, i.e., VSWR = 2 and VSWR = 3, memory effect becomes more severe. We can observe that, with the similar complexity, the EMSA model outperforms the other two models especially when the mismatch is stronger.

The performance and complexity results of NMSE and ACPR without and with DPD are shown in Table III. From the results, we can see that, when VSWR = 1, i.e., there is no standing waves in the output port of PA, MSA model and the proposed EMSA model have similar performance in both complexity and linearization performance, while the GMP model shows slightly worse NMSE performance

and higher complexity. The difference becomes much more noticeable when load mismatch appears. When VSWR = 3, the proposed DPD model achieves −36.38 dB NMSE and −45.00/−44.18 dBc ACPR with 294 FLOPs, while the GMP achieves −35.39 dB NMSE and −44.46/−43.72 dBc ACPR with more than 1000 FLOPs. For the MSA model with similar level of FLOP number, the NMSE is also 1 dB worse than the proposed EMSA model. The results show that, compared with the conventional DPD methods, the proposed EMSA model shows better linearization and has a lower complexity when the output of PA is not well matched.

D. DPD Tests on PA 2

The three DPD models with the same configurations were further tested on PA 2. The experimental spectral results of three models in the three cases of VSWR = 1, VSWR = 2, and

VSWR = 3 are shown in Fig. 10(a)–(c). Fig. 10(d)–(f) displays the AM–AM and AM–PM results of the proposed EMSA model with different VSWR values. It clearly demonstrates that the proposed EMSA model achieved a better linearization performance than the GMP and the MSA model with similar amount of FLOPs. The advantage becomes larger when the standing wave is stronger.

Table IV displays the linearization and complexity performance with VSWR = 1, VSWR = 2, and VSWR = 3. In the case without standing waves, the MSA model and the proposed EMSA model show a comparable linearization performance while maintaining lower complexity than the GMP model. The EMSA model adopting 294 FLOPs obtains −35.99 dB NMSE and −44.35/−44.44 dBc ACPR, and MSA employing 329 FLOPs achieves −36.03 dB NMSE and −43.99/−43.92 dBc ACPR, whereas GMP using 1066 FLOPs has −35.57 dB NMSE and −44.33/−44.45 dBc. Similar to what we have found on PA 1, the proposed EMSA model is demonstrated to have a better linearization performance and employs fewer FLOPs when standing waves exist at the output port of PA. With VSWR = 3, the proposed EMSA model significantly outperforms the MSA model and the GMP model. NMSE rises from −35.92 to −35.19 dB when shifting from the proposed EMSA model (294 FLOPs) to the GMP model (1066 FLOPs).

Comparing the spectrum plots of PA 1 and PA 2, taking Figs. 9(c) and 10(c) as an example, PA 2 shows stronger nonlinearity than PA 1 with standing waves, as the spectrum is very asymmetric. Still, the proposed EMSA model exhibits powerful linearization capability in this challenging situation. When VSWR = 3, the EMSA model (294 FLOPs) exceeds the GMP (1066 FLOPs) and MSA (329 FLOPs) models in NMSE by nearly 1 dB. For ACPR, the EMSA model also shows 1-dB improvement over the MSA model with similar complexity.

V. Conclusion

To improve the linearization performance of DPD models when facing the challenging situation of high VSWR, a new behavioral model has been proposed and implemented with the MSA structure. The extension of MSA model based on careful analysis of physical PA behaviors improves model performance while fulfilling the requirement of low complexity. The experimental results clearly show that the proposed EMSA model outperforms the existing models while requiring far fewer FLOPs. Therefore, the EMSA model can be a promising option for implementing cost-effective DPD solutions in 5G handsets, due to its low-cost hardware implementation and high linearization performance.

References

[1] S. C. Cripps, *RF Power Modeling for Wireless Communications*, vol. 2. Norwood, MA, USA: Artech House, 2018.

[2] L. C. Nunes, P. M. Cabral, and J. C. Pedro, "AM/AM and AM/PM distortion generation mechanisms in Si LDMOS and GaN HEMT based RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 799–809, Apr. 2008.

[3] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1880–1890, May 2017.

[4] L. Guan and A. Zhu, "Green communications: Digital predistortion for wideband RF power amplifiers," *IEEE Microw. Mag.*, vol. 15, no. 7, pp. 84–99, Nov./Dec. 2014.

[5] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.

[6] L. Ding *et al.*, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[7] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

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

[9] L. Guan and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 866–872, Apr. 2010.

[10] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 3, pp. 980–987, Mar. 2017.

[11] F. M. Barradas, T. R. Cunha, P. M. Lavrador, and J. C. Pedro, "Polynomials and LUTs in PA behavioral modeling: A fair theoretical comparison," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 12, pp. 3274–3285, Dec. 2014.

[12] L. O. Chua and A.-C. Deng, "Canonical piecewise-linear representation," *IEEE Trans. Circuits Syst.*, vol. 35, no. 1, pp. 101–111, Jan. 1988.

[13] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 737–744, Feb. 2015.

[14] W. Cao and A. Zhu, "A modified decomposed vector rotation-based behavioral model with efficient hardware implementation for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2443–2452, Jul. 2017.

[15] W. Cao, Y. Li, and A. Zhu, "Magnitude-selective affine function based digital predistorter for RF power amplifiers in 5G small-cell transmitters," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2017, pp. 1539–1541.

[16] Y. Li, W. Cao, and A. Zhu, "Instantaneous sample indexed magnitude-selective affine function-based behavioral model for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 11, pp. 5000–5010, Nov. 2018.

[17] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 4, pp. 1166–1174, Apr. 2012.

[18] Y. Li, X. Wang, J. Pang, and A. Zhu, "Boosted model tree-based behavioral modeling for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 9, pp. 3976–3988, Sep. 2021.

[19] A. Brihuega, M. Abdelaziz, L. Anttila, Y. Li, A. Zhu, and M. Valkama, "Mixture of experts approach for piecewise modeling and linearization of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 1, pp. 380–391, Jan. 2022.

[20] *User Equipment (UE) Conformance Specification: Radio Transmission and Reception*, document 38.521, 3GPP, 2011.

[21] M. Abdelaziz *et al.*, "Mobile transmitter digital predistortion: Feasibility analysis, algorithms and design exploration," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Oct. 2013, pp. 2046–2053.

[22] C. D. Presti, D. F. Kimball, and P. M. Asbeck, "Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA PA module," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 604–618, Mar. 2012.

[23] K. F. Liang *et al.*, "A quadratic-interpolated LUT-based digital predistortion technique for cellular power amplifiers," *IEEE Trans. Circuits Syst.*, vol. 61, no. 3, pp. 133–137, Mar. 2014.

[24] C. F. Goncalves, F. M. Barradas, L. C. Nunes, P. M. Cabral, and J. C. Pedro, "Quasi-load insensitive Doherty PA using supply voltage and input excitation adaptation," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 1, pp. 779–789, Jan. 2022.

[25] C. R. Chappidi, T. Sharma, and K. Sengupta, "Multi-port active load pulling for mm-wave 5G power amplifiers: Bandwidth, back-off efficiency, and VSWR tolerance," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 2998–3016, Jul. 2020.

[26] A. Scuderi, L. L. Paglia, A. Scuderi, F. Carrara, and G. Palmisano, "A VSWR-protected silicon bipolar RF power amplifier with soft-slope power control," *IEEE J. Solid-State Circuits*, vol. 40, no. 3, pp. 611–621, Mar. 2005.

[27] F. Carrara, C. D. Presti, A. Scuderi, C. Santagati, and G. Palmisano, "A methodology for fast VSWR protection implemented in a monolithic 3-W 55% PAE RF CMOS power amplifier," *IEEE J. Solid-State Circuits*, vol. 43, no. 9, pp. 2057–2066, Sep. 2008.

[28] G. D. Singh, H. M. Nemati, and L. C. N. de Vreede, "A low-loss load correction technique for self-healing power amplifiers using a modified two-tap six-port network," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 9, pp. 4069–4081, Sep. 2021.

[29] N. S. Mannem, M.-Y. Huang, T.-Y. Huang, and H. Wang, "A reconfigurable hybrid series/parallel Doherty power amplifier with antenna VSWR resilient performance for MIMO arrays," *IEEE J. Solid-State Circuits*, vol. 55, no. 12, pp. 3335–3348, Dec. 2020.

[30] H. Lyu, Y. Cao, and K. Chen, "Linearity-enhanced quasi-balanced Doherty power amplifier with mismatch resilience through series/parallel reconfiguration for massive MIMO," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 4, pp. 2319–2335, Apr. 2021.

[31] A. Kuchikulla, "Applying impedance tuning to maximize antenna performance in 5G phones," Qorvo, Greensboro, NC, USA, Tech. Rep., 2020.

[32] H. Lyu and K. Chen, "Wideband quasi-balanced Doherty power amplifier with reciprocal main/auxiliary setting and mismatch-resilient parallel/series reconfiguration," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2021, pp. 736–739.

[33] H. Zargar, A. Banai, and J. C. Pedro, "A new double input-double output complex envelope amplifier behavioral model taking into account source and load mismatch effects," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 766–774, Feb. 2015.

[34] K. Hausmair *et al.*, "Prediction of nonlinear distortion in wideband active antenna arrays," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov. 2017.

[35] F. M. Barradas, P. M. Tomé, J. M. Gomes, T. R. Cunha, P. M. Cabral, and J. C. Pedro, "Power, linearity, and efficiency prediction for MIMO arrays with antenna coupling," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 5284–5297, Dec. 2017.

[36] J. Ren, "A new digital predistortion algorithms scheme of feedback FIR cross-term memory polynomial model for short-wave power amplifier," *IEEE Access*, vol. 8, pp. 38327–38332, 2020.

[37] E. Zenteno, M. Isaksson, and P. Händel, "Output impedance mismatch effects on the linearity performance of digitally predistorted power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 754–765, Feb. 2015.

[38] A. S. Tehrani, H. Cao, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.
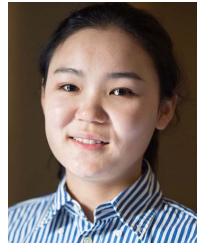
Xiaoyu Wang (Graduate Student Member, IEEE) received the B.E. degree in information engineering from Southeast University, Nanjing, China, in 2015. She is currently pursuing the Ph.D. degree at University College Dublin (UCD), Dublin, Ireland.

She is currently with the RF and Microwave Research Group, UCD. Her current research focuses on digital predistortion for RF power amplifiers, with a particular emphasis on applications to multiple-input multiple-output (MIMO) systems.
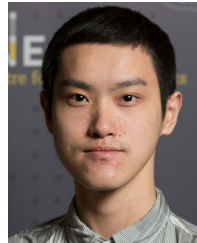
Yue Li (Member, IEEE) received the B.E. degree in information engineering from Southeast University, Nanjing, China, in 2016, and the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2020.

He is currently a Post-Doctoral Researcher with the RF and Microwave Research Group, UCD. His current research interests include behavioral modeling and digital predistortion for RF power amplifiers.

Anding Zhu (Fellow, IEEE) received the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2004.

He is currently a Professor with the School of Electrical and Electronic Engineering, UCD. His research interests include high-frequency nonlinear system modeling and device characterization techniques, high-efficiency power amplifier design, wireless transmitter architectures, digital signal processing, and nonlinear system identification algorithms.

Prof. Zhu is an elected member of MTT-S AdCom, the Chair of the Electronic Information Committee, and the Vice Chair of the Publications Committee. He is also the Chair of the MTT-S Microwave High-Power Techniques Committee. He served as a Secretary for MTT-S AdCom in 2018. He was the General Chair of the 2018 IEEE MTT-S International Microwave Workshop Series on 5G Hardware and System Technologies (IMWS-5G) and a Guest Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES on 5G Hardware and System Technologies. He is currently an Associate Editor of *IEEE Microwave Magazine* and a Track Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES.