Title	Sampling Rate Reduction for Digital Predistortion of Broadband RF Power Amplifiers
Authors(s)	Li, Yue; Wang, Xiaoyu; Zhu, Anding
Publication date	2020-03
Publication information	IEEE Transactions on Microwave Theory and Techniques, 68 (3): 1054-1064
Publisher	IEEE
Item record/more information	http://hdl.handle.net/10197/11359
Publisher's statement	© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Publisher's version (DOI)	10.1109/tmtt.2019.2944813

Sampling Rate Reduction for Digital Predistortion of Broadband RF Power Amplifiers

Yue Li, *Student Member, IEEE,* Xiaoyu Wang, *Student Member, IEEE,* and Anding Zhu, *Senior Member, IEEE*

***Abstract*—In this paper, we present a novel technique to build digital predistorters (DPD) that can linearize broadband power amplifiers (PA) using reduced sampling rates. In contrast to conventional DPDs where oversampling is necessary to avoid aliasing effect, the proposed method cancels the aliasing distortion using a sliced multi-stage cancellation scheme. A large reduction of sampling rate can be achieved in digital implementation of DPD, significantly reducing power consumption and implementation cost. Experimental results show that a DPD with a sampling rate of merely 1.5 times, instead of 5 times, signal bandwidth, can still produce satisfactory performance within the linearization bandwidth but consume only one third of power, compared with that using the conventional approaches. The proposed technique provides a promising solution for next generation 5G systems where large signal bandwidths are required.**

***Index Terms*—Aliasing effect, behavioral model, digital predistortion, linearization, power amplifiers, wireless transmitter**

I. Introduction

In next generation communication systems, e.g., 5G, wider bandwidth at higher frequency bands will be used to meet the increasing demand for higher data rates [1]. Broadband signals however may excite more severe nonlinearities of wireless transmitters, especially that induced by radio frequency (RF) power amplifiers (PA), making linearity a serious issue in system design. Digital predistortion (DPD) has been widely recognized as a highly effective approach to the linearization of RF PAs [2], [3]. Many DPD models have been developed, including memory polynomials (MP) [4], generalized memory polynomial (GMP) [5], dynamic deviation reduction (DDR) [6], decomposed vector rotation (DVR) [7], magnitude-selective affine (MSA) [8], etc. To ensure the performance of DPD models, conventional DPD methods use high sampling rates, typically 5 times signal bandwidth. Such configuration requires high clock rates of digital circuits and high speed data converters, as well as wideband transmitters and observation receiver chains.

Recent research has demonstrated that several techniques can be used to reduce the sampling rate requirement of DPD. In [9], authors proposed to constrain the bandwidth of DPD model and linearize PA within a limited bandwidth. In this way, the feedback signal can be filtered before sampling without affecting system performance. A different approach was developed in [10]–[14], where the feedback signal is also filtered before sampling but the DPD produces full-band signals. In this case, the model extraction algorithms need to apply spectral extrapolation to the loss function to achieve full-band distortion cancellation. Another method is to remove the anti-aliasing filter before sampling [15]–[18]. The aliased feedback signal is then used together with specially processed input signals to extract model coefficients. Up to date, these techniques only reduce the sampling requirement for the data converters. To generate the predistortion signal, a high clock rate is still required in the digital signal processing unit.

With increasing demands for wide signal bandwidths in the 5G systems, high speed digital signal processing in DPD will lead to high power consumption and greatly increase the overhead of base stations [19]. Unfortunately, simply reducing the sampling rate will cause aliasing effect that deteriorates performance. New methods to lower sampling rate and power consumption of DPD without sacrificing performance is thus desired. A joint in-band/out-of-band DPD [20] used multiple signal chains to form a wideband signal, which does not necessarily reduce the overall power consumption. In [21], authors targeted at sideband replica modeling and proposed to add cross-terms into existing model expressions to partially compensate for aliasing distortion at a low sampling rate. A different work [22] employed a decomposed piecewise technique where each piecewise segment is filtered before used as basis functions. Aliasing distortion is alleviated as high-order polynomials are avoided, though the piecewise operation itself may introduce additional aliasing distortion.

In this paper, a novel systematic approach to building DPD at reduced sampling rates is presented. After theoretical examination of the distortion generated by aliasing effect, a general aliasing cancellation framework is developed to eliminate its detrimental effects. The complexity of the baseline architecture is significantly reduced by applying new techniques to build memory terms and reorganize DPD basis functions. It is shown that, by employing the proposed method, DPD with a sampling rate of merely 1.5 times signal bandwidth can still produce satisfactory linearization performance. Thus, the sampling rate, and more importantly the power consumption, of DPD systems can be significantly reduced, leading to a highly energy-efficient digital linearization system for broadband 5G PAs.

The rest of this paper is organized as follows: In Section II, the outline of system architecture and the major problems caused by down-sampling are described. Section III discusses a general aliasing cancellation scheme and several complexity reduction techniques. Section IV reports the experimental results and complexity analysis, followed by a conclusion in

This work was supported in part by the Science Foundation Ireland under Grant Numbers 13/RC/2077, 16/IA/4449 and 17/NSFC/4850. (*Corresponding author: Yue Li*)

The authors are with the School of Electrical and Electronic Engineering, University College Dublin, Dublin 4, Ireland. (e-mail: yue.li1@ucdconnect.ie; xiaoyu.wang1@ucdconnect.ie; anding.zhu@ucd.ie)

Section V.

II. Sampling Rate Issue of DPD

The PA is a nonlinear device that creates spectral re-growth to the transmit signal. DPD uses an inverse model to compensate for the nonlinearity induced by the PA and thus DPD is also a nonlinear function. To effectively remove the distortion, DPD usually needs to generate a predistorted signal with 5 times bandwidth of the original transmit signal. According to Nyquist Theorem, the sampling rate of DPD must be high enough, e.g., 5 times bandwidth of the original signal if baseband complex I/Q signals are used, to avoid aliasing effect. As the bandwidth of wireless communication systems continues to increase, the sampling rate of DPD can become a serious issue in system design. For example, to process a 200-MHz signal, a 1-GHz clock is required in DPD implementation, which is not only power hungry but also challenging to design even with the most advanced CMOS technology.

A. Band-limited DPD

As discussed in [9], in practice, it may not be necessary to linearize the PA in the full bandwidth, e.g., up to 5 times bandwidth, because we may only need to remove the distortion near the center frequency band while the distortion beyond that band can be filtered by using a bandpass filter at the PA output. In [9], a band-limited DPD technique was proposed, in which a band-limiting function is inserted into the general Volterra operators in the DPD model to control the signal bandwidth under modeling, which logically transforms the general Volterra series-based model into a band-limited version. This approach allows users to arbitrarily choose the bandwidth to be linearized in the PA output according to the system requirement without sacrificing performance, making the DPD system design much more flexible and feasible.

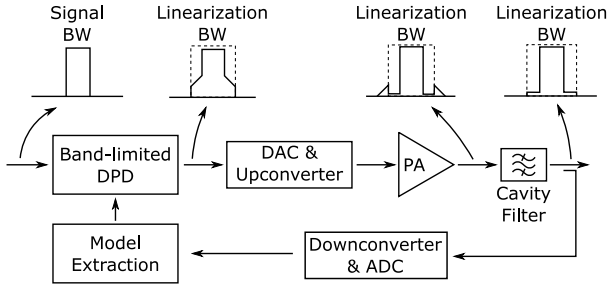

Fig. 1. System architecture of band-limited DPD.

In the band-limited DPD framework, two immediate benefits can be obtained: 1) Because the signal components beyond the linearization bandwidth are filtered, the DPD output becomes band-limited and thus we can use low sampling rate Digital to Analog Converters (DAC) when converting the DPD signal to the analog domain. For instance, as shown in Fig. 1, if the linearization bandwidth is only 1.5 times the original signal bandwidth, the sampling rate of DAC can be as low as the Nyquist rate of 1.5 rather than 5 times the original signal bandwidth. 2) Since we only focus on distortion within a specific bandwidth, we don't have to capture the signal outside that frequency range, and the sampling rate of Analog to Digital Converters (ADC) in the feedback loop can also be reduced, e.g., using a sampling rate at 1.5 rather than 5 times the original signal bandwidth. The band-limited DPD approach also releases some bandwidth requirements for the up- and down-conversion chains since a narrower signal bandwidth is required. To remove the distortion beyond the linearization bandwidth, a bandpass filter is required at the PA output.

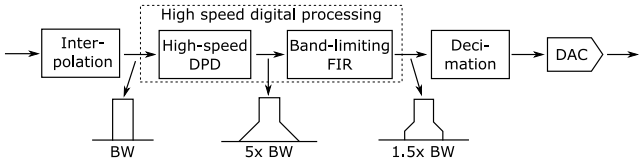

Fig. 2. Example implementation of high-speed band-limited DPD.

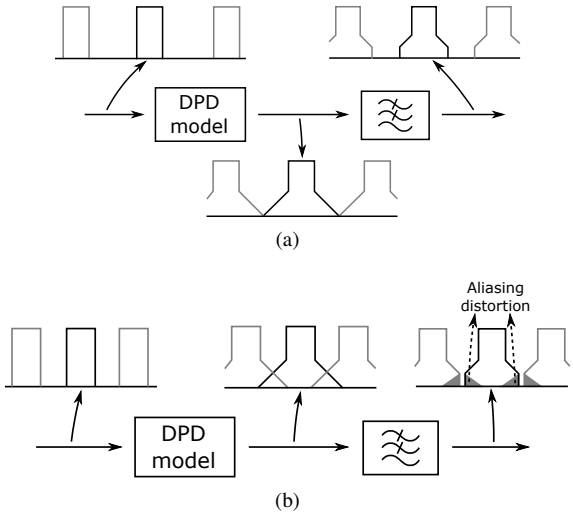

Fig. 3. Band-limited DPD processing at (a) high sampling rate and (b) low sampling rate.

B. Aliasing Effect in Digital Signal Processing

The band-limited DPD approach reduces the sampling rate of data converters. However, to generate the band-limited output signal, high speed digital signal processing must be conducted. Specifically, during the signal generation process, a high clock rate is still required inside the DPD. This is because the bandwidth reduction is achieved by inserting a digital filter after the nonlinear functions. As shown in Fig. 2, the signal bandwidth may expand to multiple times the original bandwidth during the nonlinear DPD operation. A high sampling rate must be used before filtering, otherwise alias will be introduced into the signal. In real operation, the input signal is usually interpolated to a higher sampling rate before entering the DPD model. After the nonlinear process, a band-limiting filter is applied to the signal to filter out unwanted out of band signals and then the output signal can be decimated to a lower sampling rate before entering DAC, as shown in Fig. 3(a). Otherwise, if the nonlinear function runs at low speed, aliasing effect will appear immediately after

DPD function and before the band-limiting filters. As shown in Fig. 3(b), in this case, the aliasing elements will fall into the filtered signal. Once the aliasing distortions are present, they are indistinguishable from the desired signal. These distortions will degrade the linearization performance.

III. Sampling Rate Reduced DPD

In this work, we aim to develop a new approach that enables DPD to operate at a reduced sampling rate without losing performance. Our goal is to generate an aliasing-free output with low speed digital signal processing in DPD directly, as shown in Fig. 4. To achieve this, the concept of aliasing cancellation is first introduced. A new sampling rate reduced DPD architecture is subsequently proposed by identifying the redundancy in the system and restructuring the DPD to reduce computational complexity. Please note that the sampling rate here is referred to the number of data samples per second to be processed in the digital circuit.

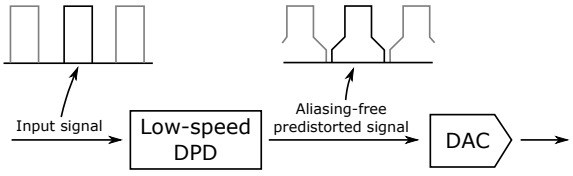

Fig. 4. Desired low speed DPD signal generation.

A. Aliasing Cancellation

To reduce the sampling rate of DPD without changing its behavior, a time-interleaved architecture as shown in Fig. 5 can be adopted. The high sampling rate DPD is transformed into a parallel structure where each path includes a copy of DPD operating at a lower sampling rate, resulting in an implementation similar to [23]. This architecture however still needs high speed interpolation before DPD and high speed filtering after DPD. The overall system complexity is also high because multiple replicated DPD blocks must be deployed. Nevertheless, this architecture provides us with a starting point to derive a new solution.

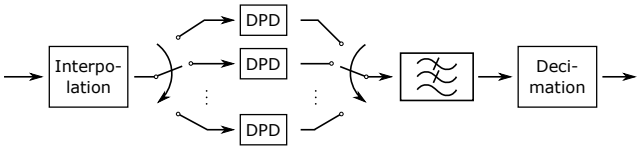

Fig. 5. Time-interleaved DPD for reducing sampling rate.

First of all, let's use a frequency domain analysis to gain some insights of this parallel system. Here, we denote the high-speed output signal of DPD as $u(n)$, and its Fourier transform as $U(j\omega)$. The time-interleaved signals are directly down-sampled versions of $u(n)$, which are $u(Dn)$, $u(Dn+1)$, etc. The Fourier transform of $u(Dn)$ can be derived as [24]:

$$\begin{aligned}\mathcal{F}\left[u(Dn)\right] &= \frac{1}{D}\sum_{p=0}^{D-1} U\left(\frac{j(\omega-2p\pi)}{D}\right) \\ &= \frac{1}{D}\left[U\left(\frac{j\omega}{D}\right) + \sum_{p=1}^{D-1} U\left(\frac{j(\omega-2p\pi)}{D}\right)\right],\end{aligned} \tag{1}$$

where $U(j\omega/D)$ is the desired signal, and other terms are frequency-shifted signals that create aliasing effect. For simplicity, we define

$$U_p(j\omega) \triangleq U\left(\frac{j(\omega-2p\pi)}{D}\right). \tag{2}$$

Thus, (1) becomes

$$\mathcal{F}\left[u(Dn)\right] = \frac{1}{D}\left[U_0(j\omega) + \sum_{p=1}^{D-1} U_p(j\omega)\right]. \tag{3}$$

If we focus on the first Nyquist zone, i.e. $0 \le \omega < 2\pi F_s/D$, the implication of (3) is clear: the down-sampled signal is linear combination of the desired signal $U_0(j\omega)$ and the aliasing distortion components.

The remaining parallel signals are simply delayed in time domain and their spectrum in the first Nyquist zone can be derived as:

$$\begin{aligned}\mathcal{F}\left[u(Dn-\Delta_i)\right] =& \frac{1}{D}\sum_{p=0}^{D-1} U\left(\frac{j(\omega-2p\pi)}{D}\right) e^{-j\Delta_i\frac{\omega-2p\pi}{D}} \\ =& \frac{e^{-\frac{j\Delta_i\omega}{D}}}{D}\sum_{p=0}^{D-1} U\left(\frac{j(\omega-2p\pi)}{D}\right) e^{\frac{2jp\Delta_i\pi}{D}} \\ =& \frac{e^{-\frac{j\Delta_i\omega}{D}}}{D}\left[U_0(j\omega) + \sum_{p=1}^{D-1} U_p(j\omega) e^{\frac{2jp\Delta_i\pi}{D}}\right].\end{aligned} \tag{4}$$

Although aliasing distortions appear in each branch, they are cancelled when the digital samples are combined in the time interleaved manner in the output. By using this parallel architecture, the clock rate of DPD operation can be lower but the output samples must be combined in high speed and then filtered by a high speed band-limiting filter that can increase power consumption. In the band-limited DPD operation, decimation is normally used after filtering to reduce the final sampling rate, which means that not all high speed samples are needed in the final output. This leads that, if we can combine the low speed data directly without time interleaving, high speed signal process can thus be avoided.

Let's re-check (4), one thing to notice is that there is an extra group delay $e^{-j\Delta_i\omega/D}$ in (4) compared to (3). If we can invert the effect of group delay via an delay adjustment filter (DAF), the resulting expression is also a linear combination of desired signal and aliasing components:

$$\mathcal{F}\left[u(Dn-\Delta_i)\right] e^{\frac{j\Delta_i\omega}{D}} = \frac{1}{D}\left[U_0(j\omega) + \sum_{p=1}^{D-1} U_p(j\omega) e^{\frac{2jp\Delta_i\pi}{D}}\right] \tag{5}$$

This means that, instead of combining the signal samples into output in the time interleaved manner, we can add them together directly after delay adjustment. The process is illustrated in Fig. 6 for the case of $D = 2$. In the upper branch, the spectra replica simply overlaps with the desired signal. In the lower branch, however, the replica has a 180-degree phase shift, so the aliasing components are actually "subtracted" from the signal. Therefore, after adjusting the group delay and summing them up, aliasing distortions will cancel out, resulting in a clean spectrum.

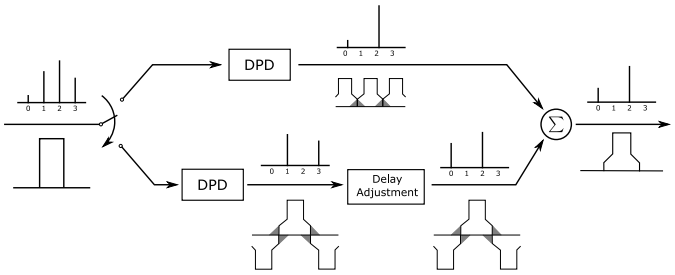

Fig. 6. Frequency and time domain demonstration of aliasing cancellation.

It is worth noting that the derivation in (4) and (5) also holds for non-uniform sampling system, i.e., Δ_i can take non-integer values. Thus, one can arbitrarily choose the delay for each parallel branch. (3) and (5) can then be written into matrix format by gathering different delayed signals:

$$\mathbf{U}_{\text{DA}} = \mathbf{H} \begin{bmatrix} U_0(j\omega) \\ U_1(j\omega) \\ \vdots \end{bmatrix}, \tag{6}$$

where

$$\mathbf{U}_{\text{DA}} = \begin{bmatrix} \mathcal{F}\left[u(Dn)\right] \\ \mathcal{F}\left[u(Dn-\Delta_1)\right] e^{j\Delta_1\omega/D} \\ \mathcal{F}\left[u(Dn-\Delta_2)\right] e^{j\Delta_2\omega/D} \\ \vdots \end{bmatrix} \tag{7}$$

$$\mathbf{H} = \frac{1}{D} \begin{bmatrix} 1 & 1 & 1 & \cdots \\ 1 & e^{j2\Delta_1\pi/D} & e^{j4\Delta_1\pi/D} & \cdots \\ 1 & e^{j2\Delta_2\pi/D} & e^{j4\Delta_2\pi/D} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{8}$$

It can be easily shown that $U_0(j\omega)$ can be reconstructed from $\mathbf{U}_{\text{DA}}$ since $\mathbf{H}$ is invertible:

$$U_0(j\omega) = \mathbf{h}_{\text{rec}} \mathbf{U}_{\text{DA}}. \tag{9}$$

where $\mathbf{h}_{\text{rec}}$ is the first row of $\mathbf{H}^{-1}$.

Herein, the principle of a general aliasing cancellation method has been developed, which can in theory perfectly recover the original high speed information by first adjusting the group delay of different parallel signals and then calculating their linear combination. A demonstration of such a system with a parallel architecture is depicted in Fig. 7, where the sampling rate of the entire system is reduced.

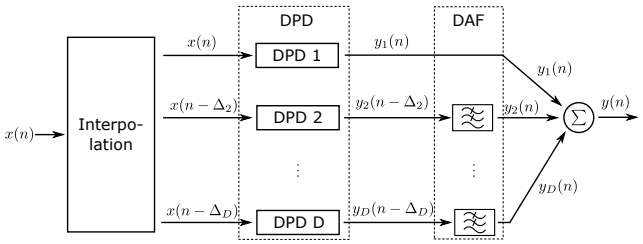

Fig. 7. Basic parallel cancellation architecture.

B. System Architecture

The parallel architecture "unrolls" the conventional band-limited DPD by converting high-speed operations to their parallel low-speed equivalence. As each branch still requires a complete DPD model, it is desirable to further reduce the hardware complexity.

1) Memory Terms Based on Delay Adjustment: The analysis and transformation start from the generation of parallel delayed input signals. Within the aliasing cancellation framework, the delays between different branches are allowed to be distributed in a non-uniform way. Thus, traditional interpolation methods need to be modified to adapt to the changes. Instead of conventional interpolation, the input signal of each branch is generated individually by a specially designed delay adjustment FIR filter aimed to reconstruct the high sampling rate input signals with corresponding delays.

The delayed signals here serve two purposes: 1) each $x(n-\Delta_i)$ is fed into the corresponding i-th DPD branch; 2) if the DPD model in use includes memory effect, the delayed signals also construct memory terms in other DPD branches. From a behavioral modeling perspective, both of them represent memory effect in the system, so delay adjustment filters actually build memory terms.

A natural question follows, can we simplify the design while maintaining its modeling capability? The answer is positive. In [25], the author proposed that for two basis functions of the same polynomial order and group delay, we may use one in model extraction and the other in predistorter, implying that the two terms have similar characteristics. That is to say, nonlinear terms may be modified without significantly affecting modeling performance if the *nonlinear order* and *group delay* are kept the same.

In the case of delay adjustment filters, no matter how the filters are designed, the nonlinear orders of memory terms are not affected by these linear filters. Therefore, the filters only need to have the correct group delay, and other factors may have little influence on the modeling performance. It suggests that delay adjustment filters don't have to perfectly reconstruct the high-speed signal, especially its amplitude. The constraints on filter design are reduced, resulting in filters with much lower orders. In this work, simple 3-tap FIR filters are enough to achieve satisfactory performance, in contrary to eighth order filters needed by full interpolation.

To design such FIR filters, the design objective is to ensure proper group delay and relatively flat amplitude response within linearization bandwidth. A simple method is to use the Fourier series method with a rectangular window, which

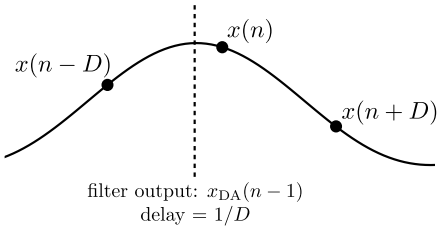

Fig. 8. FIR filter-based delay adjustment.

ensures linear phase response and constant group delay. The filter coefficients can be directly derived as

$$h_k = sinc(k - \Delta), \tag{10}$$

where k is used to index the filter coefficients.

The weighted least squares method can also be adopted, and other methods are available in [26], [27]. The desired response at different frequencies can be written as

$$\mathbf{H_d} = [H_d(\omega_1), H_d(\omega_2), ...]^T \tag{11}$$

where ω_k refers to uniformly spaced frequencies within bandwidth of interest. The desired frequency response can be represented as

$$H_d(\omega_k) = s_k e^{-j\Delta\omega_k} \tag{12}$$

where s_k is either 1 or 0.

Suppose the filter coefficients is $\boldsymbol{h} = [h_0, h_1, \cdots, h_L]$, its frequency response at frequency ω_k can be expressed as

$$H_d(\omega_k) = \sum_{n=0}^{L} h_n e^{j\omega_k(n-\frac{L}{2})}. \tag{13}$$

If we consider all frequencies of interest, (13) can be rewritten in matrix format

$$\mathbf{H_d} = \mathbf{A}\mathbf{h} \tag{14}$$

where

$$\mathbf{A} = \begin{bmatrix} e^{-j\frac{L}{2}\omega_1} & e^{j(-\frac{L}{2}+1)\omega_1} & \cdots & e^{j\frac{L}{2}\omega_1} \\ e^{-j\frac{L}{2}\omega_2} & e^{j(-\frac{L}{2}+1)\omega_2} & \cdots & e^{j\frac{L}{2}\omega_2} \\ \vdots & \vdots & & \vdots \end{bmatrix}. \tag{15}$$

And it can be solved as

$$\hat{\mathbf{h}} = (\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{H_d}. \tag{16}$$

As the filter specification is only related to the down-sampling ratio D and fraction delay Δ, the filter coefficients do not change with PA or input signal characteristics, so they can be calculated offline and viewed as constants in implementation. Under the situations where signal bandwidth needs to be dynamically reconfigured, the downsampling ratio D may be adapted for better trade-off between performance and complexity. In this case, the filter coefficients can be determined using analytical expressions like (10), thus the complexity of design process can be kept very low.

2) Model Slicing and Reorganization: In this part, we further reduce the computational complexity of the time-interleaved DPD models. To better illustrate the redundancy in the parallel structure, the original DPD model is first divided into sub-models, where each sub-model is composed of memory terms with the same memory depth. For instance, the input matrix containing all basis functions of the GMP model can be written as

$$\mathbf{X}_{\text{GMP}} = [\mathbf{X}_{\text{GMP,0}}, \mathbf{X}_{\text{GMP,1}}, \cdots], \tag{17}$$

where memory sub-models $\mathbf{X}_{GMP,i}$ only include terms with specific memory depth. Cross terms like $|\tilde{x}_{n-i-m}|^k\tilde{x}_{n-i}$ are grouped according to subscript i. The first two memory sub-models are elaborated below to better illustrate the model slicing process:

$$\mathbf{X}_{\text{GMP,0}} = \begin{bmatrix} \tilde{x}_n & |\tilde{x}_n|\tilde{x}_n & \cdots & |\tilde{x}_{n-1}|\tilde{x}_n & \cdots \\ \tilde{x}_{n-1} & |\tilde{x}_{n-1}|\tilde{x}_{n-1} & \cdots & |\tilde{x}_{n-2}|\tilde{x}_{n-1} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ \tilde{x}_2 & |\tilde{x}_2|\tilde{x}_2 & \cdots & |\tilde{x}_1|\tilde{x}_2 & \cdots \end{bmatrix} \tag{18}$$

$$\mathbf{X}_{\text{GMP,1}} = \begin{bmatrix} \tilde{x}_{n-1} & |\tilde{x}_{n-1}|\tilde{x}_{n-1} & \cdots & |\tilde{x}_{n-2}|\tilde{x}_{n-1} & \cdots \\ \tilde{x}_{n-2} & |\tilde{x}_{n-2}|\tilde{x}_{n-2} & \cdots & |\tilde{x}_{n-3}|\tilde{x}_{n-2} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ \tilde{x}_1 & |\tilde{x}_1|\tilde{x}_1 & \cdots & |\tilde{x}_0|\tilde{x}_1 & \cdots \end{bmatrix}, \tag{19}$$

where $\tilde{x}_n$ is the complex-valued baseband signal.

Based on Fig. 7, a new architecture highlighting the sliced sub-models is shown in Fig. 9. Each parallel DPD block is decomposed into multiple sub-models, and the interpolation process is replaced by the delay adjustment filters. Also, similar to Fig. 7, the parallel DPD blocks all operate at low sampling rate, so each DPD block is related to only one of D successive rows in the matrices of (17)-(19), while their memory terms are all built by the high-speed memory samples obtained from 3-tap interpolation.

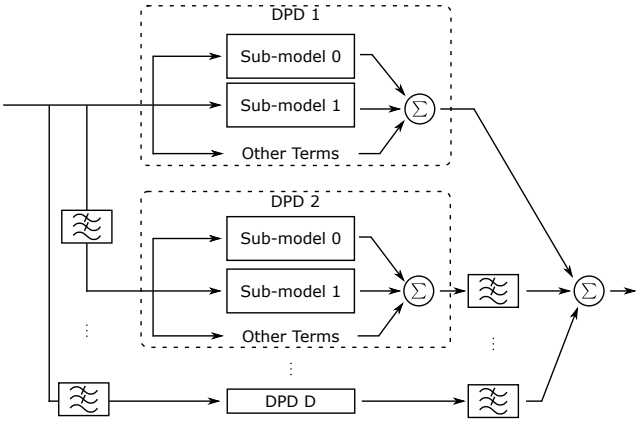

Fig. 9. Memory terms in parallel structure.

An interesting observation from (18), (19) and Fig. 9 is the similarity between the arrangement of different DPDs and different sub-models within one DPD. Take DPD 1 as an example. The input samples to build its sub-model 1 is delayed

by one sample compared with sub-model 0. In the meantime, the input samples to build the DPD 2, or more precisely sub-model 0 of DPD 2, are also delayed by one sample compared with DPD 1. It leads that the input for the sub-model 1 of DPD 1 and the sub-model 0 of DPD 2 are the same. Thus, one sub-model can share the same input samples with a different sub-model of another parallel DPD block. By organizing all sub-models with the same input samples together, Fig. 9 is changed to Fig. 10(a).

In the example above, all sub-models of a GMP model have the same nonlinear expressions, and only differ in the input samples. In this case, if two sub-models are built using the same input data, their basis functions will be exactly the same. Therefore, as shown in Fig. 10(b), sub-models that are grouped together in Fig. 10(a) can share the same nonlinear basis functions but still have their own coefficients and filter taps.

To further reduce hardware complexity, we force the multiple branches in Fig. 10(b) to share the same coefficients. The multipliers can be combined into one and built into the sub-model. The parallel delay adjustment filters can also be merged because they are all linear operators. While the complexity is reduced dramatically, the performance may be compromised because the modeling capability of memory effect is much weaker. To compensate for it, an embedded FIR filter is cascaded with the memory sub-model to better shape the frequency response of model. As both embedded and delay adjustment filters are FIR-type, they can be merged together through a convolution. Therefore, only one FIR filter is required in each path, resulting in the structure shown in Fig. 10(c). By gathering all sub-models, the time-interleaved DPD structure is finally transformed into a sliced two-stage architecture depicted in Fig. 11.

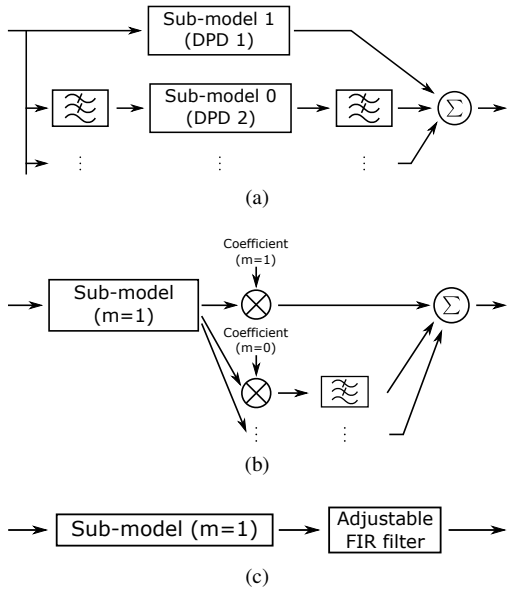

Fig. 10. Steps to reorganize sub-models.

It is worth noting that the original low-speed model is already included in the new architecture, because setting all embedded filters to 1 will reduce the proposed model to

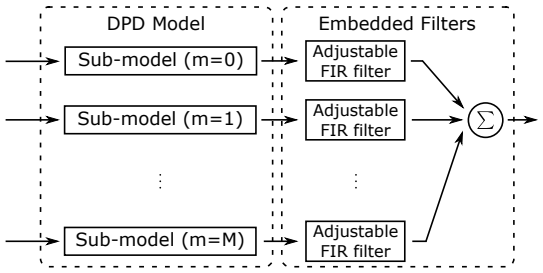

Fig. 11. Illustration of model slicing and enhancement.

conventional low-speed model. Therefore, unlike the parallel structure in previous part, proposed model integrates DPD and aliasing cancellation into the same model structure, and does not need to explicitly include a main path (the one with no delay adjustment). DPD models other than GMP can also be sliced and cascaded with embedded filters in a similar way without significantly affecting the cancellation performance. Since the aliasing distortion comes from frequencies far from in-band, the power of aliasing distortion is typically much smaller than the nonlinearity generated by the model, so aliasing cancellation may have a much lower accuracy requirement than DPD modeling. As the embedded filters are only responsible for aliasing cancellation, the approximation with a parallel two-stage model is enough to produce satisfactory performance.

3) Complete System Structure: The overall system architecture is summarized in Fig. 12, which includes memory interpolation and sliced parallel compensation, both operating at a reduced sampling rate. A low-order band-limiting filter can be optionally adopted to further constrain the linearization bandwidth so that better performance can be achieved.

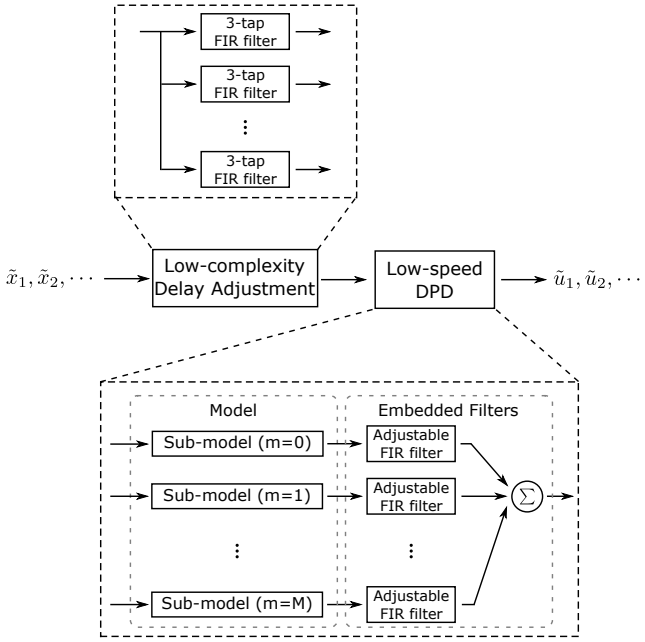

Fig. 12. Complete architecture of the proposed DPD.

Compared with conventional DPD systems, the proposed architecture significantly reduces computational complexity

and the related power consumption. The original interpolation operation is replaced by a low-complexity delay adjustment block operating at low clock rate. The core DPD function can also run at a lower sampling rate. The significant saving on power consumption from the reduced sampling rate makes the additional complexity from embedded filters negligible. The original high-order band-limiting filter becomes an optional low-order filter, and can be completely removed to keep minimal hardware complexity. Moreover, as all blocks run at the same reduced sampling rate, the original high-rate system is transformed into a low-rate system, so simpler hardware design rules and procedures can be adopted.

The proposed method also favors a different design methodology. The conventional thinking may first determine the sampling rate of the original high-speed system and then choose an appropriate down-sampling ratio. Instead, we start the design process by determining the reduced sampling rate, which reflects the clock rate of digital circuits and is usually subject to hardware and power constraints. Once the "real" sampling rate is set, the amount of aliasing distortion generated by a given DPD model is determined. Thereby, the next parameters to decide are the number of parallel delay adjustment filters and the length of each embedded filter. We view these parameters as a trade-off between hardware complexity and the performance of aliasing cancellation. By employing more complex filter configurations, the proposed method can access to more high sampling rate information and provide better aliasing cancellation performance. The final step is to choose the delay for each parallel branch. Uniform delays are straightforward, but it is also possible to manually set different delays. Additional design space can be exploited to fully optimize the expressive power of memory terms, and the non-uniformity may also lead to suppression of aliasing effect.

C. Model Extraction

As a consequence of the modified model structure, the model cannot be directly extracted with least squares (LS). Herein, an iterative two-step methodology is adopted: The coefficients of original model, $\mathbf{C}$, is first extracted, followed by the coefficients in the embedded FIR filters, $\mathbf{C}_h$.

Before model extraction, a set of input and output data samples from PA measurements are captured. To facilitate the discussion, we express the DPD process in matrix format. As the model includes multiple paths, for convenience, the two sets of coefficients $\mathbf{C}$ and $\mathbf{C}_h$ are both arranged according to the memory depth:

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_0 \\ \mathbf{C}_1 \\ \vdots \\ \mathbf{C}_M \end{bmatrix} \tag{20}$$

$$\mathbf{C}_h = \begin{bmatrix} \mathbf{C}_{h,0} \\ \mathbf{C}_{h,1} \\ \vdots \\ \mathbf{C}_{h,M} \end{bmatrix} \tag{21}$$

where $\mathbf{C}_m$ and $\mathbf{C}_{h,m}$ are the coefficients of original model and embedded filter of memory depth m, respectively.

In step one, we write the matrix formula by keeping $\mathbf{C}_h$ constant:

$$\mathbf{U} = \mathbf{X}\mathbf{C} \tag{22}$$

where $\mathbf{U}$ is the output vector of DPD, and

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_0 * \mathbf{C}_{h,0} & \mathbf{X}_1 * \mathbf{C}_{h,1} & \cdots & \mathbf{X}_M * \mathbf{C}_{h,M} \end{bmatrix} \tag{23}$$

where $\mathbf{X}_m$ is the input matrix within m-th memory sub-model. $*$ represents convolution operation. The formulation of $\mathbf{X}_m$ is the same as the model slicing operation shown in (17) to (19).

The original model can thus be extracted using LS method as

$$\mathbf{C} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{U}. \tag{24}$$

Note that in the first iteration where the value of $\mathbf{C}_h$ is not available, the convolution in (23) should be ignored.

The next step is to extract the embedded filters. We first calculate the output of every memory sub-model

$$\mathbf{u}_m = \mathbf{X}_m\mathbf{C}_m. \tag{25}$$

As the sub-models are filtered individually, the adjustable filter in each path can be written into the following matrix representation:

$$\mathbf{u}_{h,m} = \mathbf{U}_m\mathbf{C}_{h,m} \tag{26}$$

where

$$\mathbf{U}_m = \begin{bmatrix} \mathbf{u}_m(n) & 0 & 0 & \cdots & 0 \\ \mathbf{u}_m(n-1) & \mathbf{u}_m(n) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_m(n-l) & \mathbf{u}_m(n-l+1) & \mathbf{u}_m(n-l+2) & \cdots & \mathbf{u}_m(n) \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \tag{27}$$

and l is the length of adjustable filter.

Therefore, the predistorted signal $\mathbf{U}$ can be expressed in terms of $\mathbf{C}_h$ by

$$\mathbf{U} = \mathbf{U}_h\mathbf{C}_h \tag{28}$$

where

$$\mathbf{U}_h = \left[\mathbf{U}_0, \mathbf{U}_1, \cdots, \mathbf{U}_M\right]. \tag{29}$$

The embedded FIR filter can thus be extracted using LS method as

$$\mathbf{C}_h = \left(\mathbf{U}_h^H\mathbf{U}_h\right)^{-1}\mathbf{U}_h^H\mathbf{U}. \tag{30}$$

The full model can be accurately extracted by iterating between the two steps. As both steps employ the LS method, fast and stable convergence is still ensured. Typically, 2-3 iterations are enough to obtain satisfactory results. Also, as the number of coefficients in adjustable FIR filters is much smaller than that in the main DPD model, there is little additional complexity in model extraction for the proposed method.

IV. Results

A. Experimental Results

To validate the model performance, a test platform was set up, as shown in Fig. 13, which includes PC, signal generator, driver amplifier, PA, attenuator and spectrum analyzer. The PA under test is an in-house designed broadband Doherty power amplifier operating at 3.75 GHz with 31.3 dBm output power. The excitation input signal is a 100 MHz LTE signals with 6.5 dB peak-to-average power ratio (PAPR). Recorded I/Q input and output samples were time aligned and normalized before training the model. The model extraction and predistorted signal generation were performed in MATLAB. GMP model and indirect learning method were employed in the test, though other models and model extraction algorithms could also be used. The GMP model had memory depth 7, polynomial order 7 and cross term memory delay 2, resulting in 248 coefficients. The proposed method divided the GMP model into 8 sub-models (determined by the memory depth). The adjustable filter in the proposed method has length 9 for 150 MHz sampling rate tests and length 3 for 200 MHz sampling rate tests.

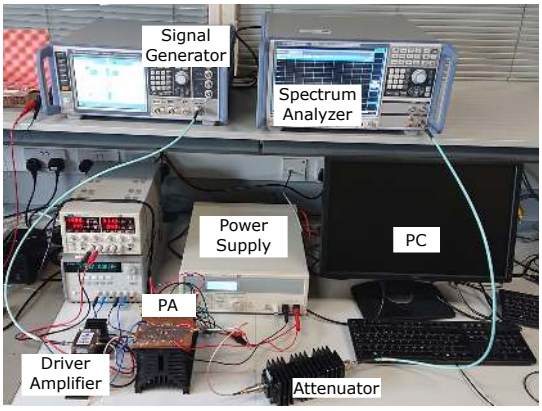

Fig. 13. The photograph of the DPD test bench.

First, no additional band-limiting filters were used. To produce strong aliasing distortions, sampling rates for conventional low-speed and proposed methods were set to 150 MSPS, while the conventional high-speed case used 600 MSPS. In real band-limited test, the PA output is supposed to be filtered and thus the un-linearized out of band distortion should be removed. In our test, the RF filter was not used in order to make better comparisons between different cases, e.g., keep the sideband distortion of the proposed case in the spectrum plots. Because of the limited linearization bandwidth, adjacent channel power ratio (ACPR) was measured at ± 55 MHz offset and the measurement bandwidth was set to 9 MHz. The spectrum results using GMP model are depicted in Fig. 14. It is shown that the proposed DPD architecture achieved around 10 dB better ACPR over the conventional low sampling rate DPD in the test. AM-AM and AM-PM results using the proposed DPD are shown in Fig. 15.

A summary of more experimental results using various sampling rates in the tests are reported in Table I. Normalized mean squared error (NMSE) performance was measured after applying the band-limiting filter to remove the sideband distortions. It is observed that the proposed method exhibits robust performance across different sampling rate settings, while the performance of conventional method can degrade quickly with reduction in sampling rate. Compared with conventional high sampling rate DPD, proposed method produces a similar near-band linearization results with a much lower sampling rate. Though it needs more coefficients, the overall computational complexity is significantly reduced, thanks to its low sampling rate.

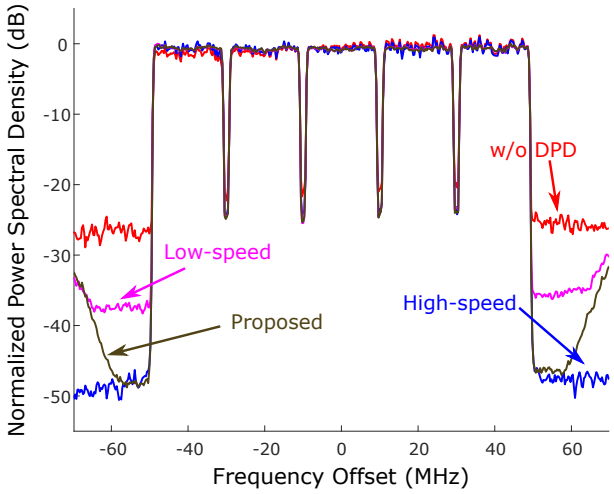

Fig. 14. Output spectra comparison for 100 MHz LTE signal using GMP model.

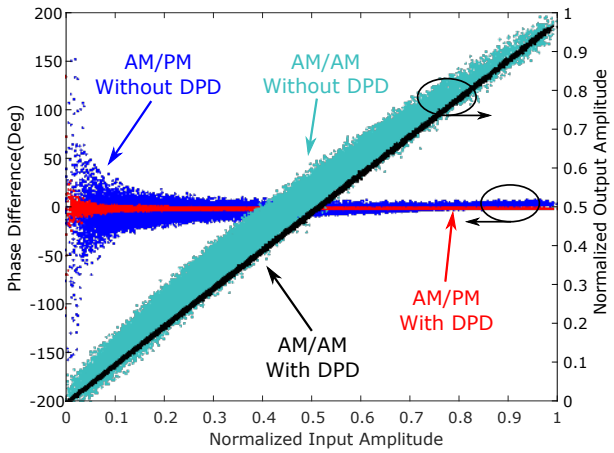

Fig. 15. AM/AM and AM/PM plots with and without proposed DPD.

TABLE I
Result Summary without Band-limiting Filter

	Sampling Rate (MSPS)	No. of Coefficients	ACPR (dBc) ($\pm$55MHz)	NMSE (dB)
w/o DPD	N/A	N/A	-26.5/-24.7	-20.6
Conventional	600	248	-46.7/-46.1	-41.3
	200	248	-43.5/-41.4	-38.6
	150	248	-36.5/-34.7	-33.7
Proposed	200	320	-47.0/-46.8	-40.7
	150	272	-47.1/-45.4	-42.6

More tests were performed after employing band-limiting filters. The band-limiting filters are FIR filters placed after DPD model. They further constrain the bandwidth of pre-distorted signal and force DPD to concentrate on narrower bandwidth, so better performance can be achieved at the expense of reduced linearization bandwidth. GMP model with the same configuration as the previous test was used. The results depicted in Fig. 16 had a similar setting as in Fig. 14, despite the use of a 65 MHz band-limiting filter. Proposed method achieved 9 dB better ACPR than conventional DPD of the same sampling rate. AM-AM and AM-PM results using the proposed DPD are shown in Fig. 17. More experimental results using different band-limiting filter settings are reported in Table II.

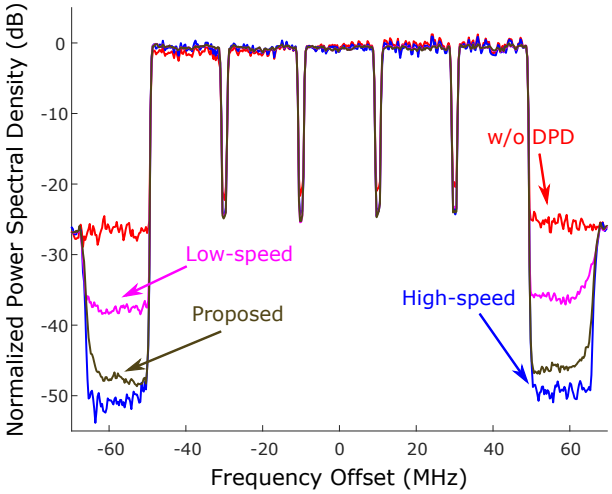

Fig. 16. Output spectra comparison with 130MHz band-limiting filter and 4 times down-sampling.

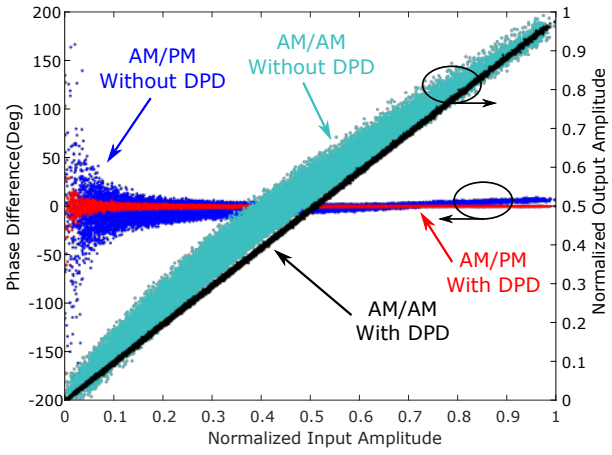

Fig. 17. AM/AM and AM/PM plots with and without proposed DPD using 65MHz band-limiting filter.

B. Complexity and Power Consumption Comparison

The complexity of DPD block using conventional and proposed configurations are compared in Table III. The same GMP model used for experimental test was employed. Since multipliers take up most of the hardware resources, hardware complexity was represented by the number of real multipliers. Another important metric, the required hardware performance, is typically measured by the number of operations per second [19]. In this work, this computational load was measured by the required real multiplication operations per second (Mult/s). Despite a larger number of coefficients, the comparison indicates that the total computational load was reduced by 65%, which is mainly because a lower sampling rate is used in the digital signal process in the proposed solution. The complexity for LS-based model extraction is given in Table IV. 20,000 I/Q samples were used for model extraction complexity comparison, showing that our method requires slightly more computation due to the two-step model extraction strategy.

TABLE II
RESULT SUMMARY WITH BAND-LIMITING FILTER

	Sampling Rate (MSPS)	Filter Cut-off Bandwidth (MHz)	ACPR (dBc) (±55MHz)	NMSE (dB)
w/o DPD	N/A	N/A	-26.5/-24.7	-20.0
Conventional	600	80	-49.0/-49.0	-43.6
	600	65	-49.0/-48.2	-43.3
	200	80	-44.2/-42.6	-38.3
	200	65	-44.1/-43.0	-39.4
	150	65	-36.8/-35.2	-34.1
Proposed	200	80	-49.1/-48.1	-43.5
	200	65	-49.9/-47.9	-41.5
	150	65	-46.9/-45.2	-42.9

TABLE III
DPD COMPLEXITY COMPARISON

	Conventional	Proposed
No. of Coefficients	248	320
Sampling Rate (MSPS)	600	150
No. of Multipliers	528	744
Computational Load (Mult/second)	316,800	111,600

TABLE IV
MODEL EXTRACTION COMPLEXITY COMPARISON

	Conventional	Proposed
No. of Coefficients	248	320
No. of Real Multiplication	642,692,992	698,506,240
Relative Complexity	100%	108.7%

The hardware power consumption of DPD block is estimated using Xilinx Power Estimator [28]. The resource utilization is estimated by counting the required resources for the IP cores of the computational components, e.g. adders, multipliers, complex multipliers, etc. The FPGA board employed was Virtex-7 XC7VX485T. Each I/Q sample of both input and output was kept as 32-bit data, where real and imaginary part each had 16 bits. A 50% toggle rate is assumed for power estimation. The FPGA resource utilization and power consumption are estimated and reported in Table V, showing the proposed DPD consumes merely 37% dynamic power of conventional DPD methods.

It is worth noting that the complexity comparison and power estimation above are indicative and aims to show the relative number/trend only. The absolute numbers and power consumption depend on the specific model used and the actual circuit implementation in real applications. Furthermore, the referred signal processing speed in Table V is defined as the number of data samples needed to be processed per second, which may not always be the same as the FPGA clock rate. For example, the 600-MHz design may be implemented using a polyphase configuration to reduce the clock rate, which may also result in slightly different power consumption [23].

TABLE V
FPGA UTILIZATION AND POWER ESTIMATION

	Conventional	Proposed
Processing Speed (MHz)	600	150
Slice LUT	9161	12905
Slice Register	9613	17821
DSP	528	744
Dynamic Power (W)	5.110	1.875

V. CONCLUSION

A novel DPD architecture has been proposed to linearize broadband RF PAs with reduced sampling rate based on a new aliasing cancellation scheme. As the power consumption of DPD may become a main obstacle for future DPD applications, the dynamic power consumption should be considered a more important factor than digital hardware resource utilization. Under this context, the proposed sliced multi-stage cancellation technique demonstrates reduced clock rate while only slightly increasing the complexity, such that a dramatic reduction in power consumption is achieved. Experimental tests have also validated the linearization performance of the proposed method, making it a feasible solution to the linearization of wideband signals in 5G.

REFERENCES

[1] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, Jun. 2017.

[2] F.-L. Luo, *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*. Cambridge, UK: Cambridge University Press, 2011.

[3] 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. 2014.

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

[5] 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, no. 10, pp. 3852–3860, Oct. 2006.

[6] 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.

[7] 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.

[8] 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.

[9] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited Volterra series-based digital predistortion for wideband RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 12, pp. 4198–4208, Dec. 2012.

[10] L. Guan, C. Yu, and A. Zhu, "Bandwidth-constrained least squares-based model extraction for band-limited digital predistortion of RF power amplifiers," in *Integrated Nonlinear Microwave and Millimetre-Wave Circuits (INMMIC), 2012 Workshop On*, IEEE, 2012, pp. 1–3.

[11] Y. Liu, J. J. Yan, H.-T. Dabag, and P. M. Asbeck, "Novel technique for wideband digital predistortion of power amplifiers with an under-sampling ADC," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2604–2617, Nov. 2014.

[12] Y. Ma, Y. Yamao, Y. Akaiwa, and K. Ishibashi, "Wideband digital predistortion using spectral extrapolation of band-limited feedback signal," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 7, pp. 2088–2097, Jul. 2014.

[13] Y. Liu, W. Pan, S. Shao, and Y. Tang, "A general digital predistortion architecture using constrained feedback bandwidth for wideband power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 5, pp. 1544–1555, May 2015.

[14] R. N. Braithwaite, "Closed-loop digital predistortion (DPD) using an observation path with limited bandwidth," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 726–736, Feb. 2015.

[15] H. Koeppl and P. Singerl, "An efficient scheme for nonlinear modeling and predistortion in mixed-signal systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 12, pp. 1368–1372, Dec. 2006.

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

[17] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low feedback sampling rate digital predistortion for wideband wireless transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 11, pp. 3528–3539, 2016.

[18] Z. Wang, L. Guan, and R. Farrell, "Undersampling observation-based compact digital predistortion for single-chain multiband and wideband direct-to-RF transmitter," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 5274–5283, Dec. 2017.

[19] 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.

[20] C. Yu, N. Yang, Q. Lu, J. Cai, and X. Zhu, "High-precision joint in-band/out-of-band distortion compensation scheme for wideband RF power amplifier linearization," *IEEE Microw. Compon. Lett.*, vol. 28, no. 11, pp. 1044–1046, Nov. 2018.

[21] W. Cao, Y. Li, and A. Zhu, "Digital suppression of transmitter leakage in FDD RF transceivers: Aliasing elimination and model selection," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1500–1511, Mar. 2018.

[22] C. Yu, Q. Lu, H. Sun, X. Wu, and X.-W. Zhu, "Digital predistortion of ultra-broadband mmWave power amplifiers with limited Tx/feedback loop/baseband bandwidth," *Wirel. Commun. Mob. Comput.*, 2018.

[23] L. Xu, T. Tran, and M. Fitton, "Designing polyphase DPD solutions with 28-nm FPGAs," Altera Corporation, San Jose, CA, White Paper WP-01171-1.0, Jan. 2012.

[24] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: A tutorial," *Proc. IEEE*, vol. 78, no. 1, pp. 56–93, Jan. 1990.

[25] R. N. Braithwaite, "Digital predistortion of an RF power amplifier using a reduced Volterra series model with a memory polynomial estimator," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3613–3623, Oct. 2017.

[26] T. W. Parks and C. S. Burrus, *Digital Filter Design*. Wiley-Interscience, 1987.

[27] T. I. Laakso, V. Valimaki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay [FIR/all pass filters design]," *IEEE Signal Process. Mag.*, vol. 13, no. 1, pp. 30–60, Jan. 1996.

[28] Xilinx, "Xilinx Power Estimator User Guide," 2017. [Online]. Available: https://www.xilinx.com/support/documentation/sw_manuals/xilinx2017_1/ug440-xilinx-power-estimator.pdf.