Compact Digital Predistortion for Multi-band and Wide-band RF Transmitters

by

Ziming Wang

Supervisor: Professor Ronan Farrell

Electrical Engineering Department

A thesis submitted to National University of Ireland Maynooth
in fulfillment of the requirements for the degree of
Doctor of Philosophy

2018

Acknowledgements

During my PhD studies, I feel very lucky to meet so many nice and talented people, who have supported me with great patience. I owe my deepest gratitude to my advisor, Prof Ronan Farrell, for providing me with the opportunity of doing my doctorate. His prospective outlook and insightful comments, constructive criticism, and friendly approach helped me a lot to guide me and enrich my ideas. Without his support, encouragement, fruitful discussions and guidance, this work would never have been possible.

I would like to thank especially Dr Lei Guan, members of technical staff in Bell Lab Dublin, for helping me throughout my internship. I also wish to express my sincere thanks to the faculty and staff of Electronic Engineering Department for providing me help and cooperation during this research work. Thanks to Jim for the help of hardware design, and Joanna for her great patience and ability to organise the office and create the pleasant atmosphere.

Special thanks to Dr Sarah Adel Ibrahim who brings so much fun to my PhD life. The daily discussion, papers review and encouragement are essential to this PhD work. Thanks to Dr John Dooley and all the PhD students in the lab for your supporting, understanding and making me miss the life in the lab.

I am extremely grateful to my Chinese friends in Maynooth. Your friendship made my days in Maynooth.

Abstract

This thesis is focusing on developing a compact digital predistortion (DPD) system which costs less DPD added power consumptions. It explores a new theory and techniques to relieve the requirement of the number of training samples and the sampling-rate of feedback ADCs in DPD systems. A new theory about the information carried by training samples is introduced. It connects the generalized error of the DPD estimation algorithm with the statistical properties of modulated signals. Secondly, based on the proposed theory, this work introduces a compressed sample selection method to reduce the number of training samples by only selecting the minimal samples which satisfy the foreknown probability information. The number of training samples and complex multiplication operations required for coefficients estimation can be reduced by more than ten times without additional calculation resource. Thirdly, based on the proposed theory, this thesis proves that theoretically a DPD system using memory polynomial based behavioural modes and least-square (LS) based algorithms can be performed with any sampling-rate of feedback samples. The principle, implementation and practical concerns of the undersampling DPD which uses lower sampling-rate ADC are then introduced. Finally, the observation bandwidth of DPD systems can be extended by the proposed multi-rate track-and-hold circuits with the associated algorithm. By addressing several parameters of ADC and corresponding DPD algorithm, multi-GHz observation bandwidth using only a 61.44MHz ADC is achieved, and demonstrated the satisfactory linearization performance of multi-band and continued wideband RF transmitter applications via extensive experimental tests.

Publications

1. **Z. Wang**, L. Guan and R. Farrell, "Undersampling Observation-Based Compact Digital Predistortion for Single-Chain Multiband and Wideband Direct-to-RF Transmitter," in IEEE Transactions on Microwave Theory and Techniques, 2017.

2. **Z. Wang**, L. Guan, and R. Farrell, Compact undersampled digital predistortion for flexible single-chain multi-band rf transmitter, in IEEE MTT-S International Microwave Symposium 2017 (IMS2017), 2017.

3. **Z. Wang**, S. Ibrahim, H. Su, and R. Farrell, Generalised digital predistortion of rf power amplifiers with low-rate feedback signal, in 2016 46th European Microwave Conference (EuMC), Oct 2016.

4. **Z. Wang**, J. Dooley, K. Finnerty, and R. Farrell, Selection of compressed training data for rf power amplifier behavioral modeling, in Proceedings of the 10th European Microwave Integrated Circuits Conference, European Microwave Association, 2015.

5. **Z. Wang**, J. Dooley, K. Finnerty, and R. Farrell, A low-rate identification scheme of high power amplifiers, in Proceedings of the 17th Research Colloquium on Communications and Radio Science, Ireland, 2014.

6. S. A. Ibrahim, **Z. Wang**, and R. Farrell, An mems phase shifter with high power handling for electronic beam tilt in base station antennas, IEEE Microwave and Wireless Components Letters, 2017.

7. S. Hesami, J. Dooley, **Z. Wang** and R. Farrell, "Digital predistorter in crosstalk compensation of MIMO transmitters," 2016 27th Irish Signals and Systems Conference (ISSC), Londonderry, 2016

8. S. A. Ibrahim, H. Su, **Z. Wang** and R. Farrell, A new rf interference cancellation using a novel 3-pole bandstop resonator designed using tapped crlh t-line model, in Microwave Conference (EuMC), 2016 46th European, pp. 108111, IEEE, 2016.

9. H. Su, **Z. Wang**, S. Ibrahim, and R. Farrell, A digitally assisted analog cancellation system at rf frequencies for improving the isolation performance of a ceramic duplexer, in Microwave Conference (EuMC), 2016 46th European, IEEE, 2016.

10. H. Su, **Z. Wang** and R. Farrell, "Compressed-sampling-based behavioural modelling technique for wideband RF transmitter leakage cancellation system," 13th International Conference on Synthesis, Modeling, Analysis and Simulation Methods (SMACD), Lisbon, 2016

Abbreviations

DPD digital predistortion

LS least-square

RF radio frequency

PA power amplifiers

5G 5th generation

ADC analog-to-digital converter

LTE-A LTE-advanced

MIMO multiple-input-multiple-output

LOs local oscillators

PDF probability distribution function

PAPR peak-average-power-ratio

PAE power added efficiency

ITU International Telecommunication Union

ET Envelope Tracking

CW continuous wave

P1dB 1 dB compression point

AM-AM amplitude-to-amplitude

AM-PM amplitude-to-phase

IMD intermodulation distortion

DAC digital-to-analog converter

LTI linear time-invariant

GMP generalized memory polynomial

IDLA indirect learning structure

DLA direct learning structure

RLS recursive least square

LMS least mean square

NMSE normalized mean square error

EVM error vector magnitude

ACPR adjacent channel power ratio

ACLR adjacent channel leakage ratio

THA track-and-hold amplifier

IF intermediate-frequency

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Wireless communications have already shaped our daily life by providing seamless radio access worldwide. According to the latest global ICT data released in 2017 [5], mobile broadband subscriptions have grown more than 20 percent annually in the last five years and are expected to reach 4.3 billion globally by the end of 2017. The increasing demand for more subscribers brings challenges in developing the next-generation wireless radio access devices. The expectation of service of wireless communications is inevitably high, and any loss in signal quality is treated as a source of concern [6]. On the other hand, the subscriber needs cheaper bundles with the higher data rate, while operators need more efficient solutions with the wider bandwidth.

As a critical physical unit between subscribers and operators, the wireless base station is the most expensive element in mobile networks. It is also the dominant power consumption compared to the other components in the mobile network. Meanwhile, over 60% of the power consumption is wasted by radio fre-

quency (RF) power amplifiers (PAs) in the form of heat rather than transmitting data [7]. In wireless communication systems, the coverage area of the base station depends on the signal power and the operating frequency. To guarantee the correct reception of RF signal at the receiver, RF PAs are employed at the transmitter-end to amplify the signal power, which is the main consumer of the power. To maintain the overall cost of operation, the PAs are operated at high-efficiency levels, in compression, close to saturation [8]. In this region, nonlinear distortions generated by PA, which is treated as in-band distortion and spectral leakage into the adjacent channels, are heavily regularized. To meet the linearity and high-efficiency requirements, it is necessary to operate the PAs at a high-efficiency level, and compensate for the nonlinear distortions by linearization techniques. The digital predistortion (DPD) is one of the linearization techniques, which has been widely applied in modern wireless communication systems because of adaptation possibilities, high linearization performance and low implementation effort.

During the past decade, DPD techniques have proven to be one of the most effective solutions to achieve the adequate linearity of RF PAs [7]. However, several challenges of DPD techniques for the forthcoming 5th generation (5G) wireless network attract attention in both academic and industry. The DPD characteristics are modeled by the nonlinear operators, which normally occupy multiple times the input bandwidth. For example, the 5th-order nonlinear operator occupies five times the input bandwidth. In order to realize accurate model extraction, the bandwidth of the PA output will also keep multiple times the input bandwidth to capture all major nonlinear distortion in the feedback path. As a result, the high-accuracy analog-to-digital converter (ADC) becomes the limiting factor for wideband DPDs and multi-band DPD. For example, multi-carrier UMTS systems support up to 60-MHz transmission bandwidth,

LTE-advanced (LTE-A) systems can utilize up to 100-MHz bandwidth using carrier aggregation techniques and to fully exploit the advantage of mmWave technique, ultra-broadband modulated signals will be employed to enable high-speed transmission. Ultrahigh-speed and high-resolution data acquisition and DSP units to properly handle those wideband signals brings the extra-cost and design complexity. Moreover, as for the traditional multi-band DPD, it is not practical to sample the signals in dual bands as a continuous band signal, due to the sampling-rate requirements for ADCs. Therefore, advanced multi-band behavioural models are proposed leading to increasing of the number of the model coefficients. It is a big challenge to develop a proper DPD technique with reasonable linearization performance under the reduced sampling-rate feedback ADC. On the other hand, since DPD algorithms are normally performed in a baseband digital processor, the extra power will be consumed for computing DPD parameters. For the low-power small cells, mobile handsets and massive multiple-input-multiple-output (MIMO) systems, the power consumed by the DPD unit cannot be ignored [7]. Thus, wideband DPD techniques requiring less computation complexity and less hardware cost are in high demand for the next generation wireless communication systems.

1.2 Thesis Contribution and Outline

This thesis aims to research the theory and techniques of DPD systems with reduced complexity in algorithm and hardware. By investigating the mechanism of coefficients calculation of memory polynomial based models, this thesis proposes a new perspective to connect the information of training samples with the coefficients calculation of memory polynomial models. Based on this thesis, the number of training samples can be reduced which leads to a significant reduction

of the calculation complexity using least-square based algorithms. On the other hand, the thesis proves that theoretically a memory polynomial based DPD can be performed using any sampling-rate of feedback signals with the same linearization performance. Consequently a generalized undersampling DPD is proposed in order to illuminate the principle, implement and practical concerns of the DPD system using undersampling ADCs (in this thesis, the term undersampling indicates the sampling-rate is lower than the Nyquist sampling-rate). Moreover, a single-ADC based DPD sampling receiver architecture which uses only one low-rate ADC to replace conventional DPD feedback paths that usually include mixers or analog demodulators, local oscillators (LOs), and middle-to-high range ADCs is introduced. The proposed single-ADC DPD system is designed to cover 4 GHz instantaneous bandwidth with only one ADC running at 61.44 MSPS. In terms of multi-band applications, only single feedback paths and single-input behavioural models are required. As a result, for continued wideband and multi-band DPD systems, the hardware cost of feedback paths can be dramatically reduced. The compact DPD in this thesis which uses less number of training samples and lower feedback sampling-rate enables a promising efficient DPD-enabled multi-band and wideband RF transmitter solution.

Chapter 2 covers system parameters of RF PA as well as PA memory effects. Behavioural modeling concepts including popular behavioural models, coefficients estimation procedure and algorithms, in-band and out-band linearization performance metrics are also introduced. At the end, a literature review for the wideband DPD which requires sampling-rate reduction techniques, and the multi-band DPD which requires multiple feedback paths and advanced MIMO behavioural models are given.

Chapter 3 introduces the theory of information carried by modulated signals for the LS based nonlinear system modeling. It defines the information of the

training samples used for the coefficients estimation is to carry a joint probability distribution function (PDF). This joint PDF determines the generalization error of the coefficients estimation. At the end, the information carried by training samples is connected with the peak-average-power-ratio (PAPR) and the bandwidth of the signal which both can be predetermined before the signal transmitting.

Chapter 4 presents an algorithm which uses the probability information of the input signal to inform the selection of a compressed training dataset for RF PA behavioural model extraction. The proposed algorithm can dramatically reduce the number of training samples. The accuracy of this algorithm is experimentally validated by extraction of behavioural models using a large dataset of consecutive samples and a reduced training dataset determined using the proposed algorithm. A noticeable reduction in computational complexity and faster execution time is achieved with the new approach.

Chapter 5 firstly presents the completed undersampling DPD theory. The detailed system architecture of the undersampling DPD is also introduced with new comprehensive practical concerns. Finally, both proposed theory and implementation are verified by comparable experimental tests using varying signals captured from different sampling-rates.

Chapter 6 introduces a novel compact undersampling digital predistortion-based PA linearization solution (including both hardware architecture and DPD algorithm) for single-chain wideband and multi-band RF transmitter. Particularly, it proposes a single-ADC based DPD sampling receiver architecture, i.e., using only one ADC with certain analog filters to replace conventional DPD feedback paths that usually include mixers or analog demodulators, LO, and middle-to-high range ADCs. By addressing several parameters of ADC and corresponding

DPD algorithm, multi-GHz DPD bandwidth coverage is achieved, and demonstrates the satisfactory linearization performance of multi-band and wideband RF transmitter applications via extensive experimental tests.

Lastly, Chapter 7 concludes the thesis by summarizing the contributions and the achieved results. Future works and possible improvement are discussed.

Chapter 2

Wideband and Multi-band Digital Predistortion

This chapter gives a literature review of wideband and multiband DPD techniques. It starts with the introduction of the RF PA (PA) which is transmitters' most challenging block. Next, it describes the concepts of DPD techniques, and latest researches in the area of wideband and multiband DPD are introduced.

2.1 PA

2.1.1 Introduction

As a key physical connection between subscribers and operators, the wireless base station is the most costly unit in cellular networks. Meanwhile, it also consumes the majority of power compared to other units in the network [9], as shown in Fig. 2.1. As shown in Fig. 2.2, a typical high-efficient wireless transmitter comprises three essential blocks: the baseband signal processor,

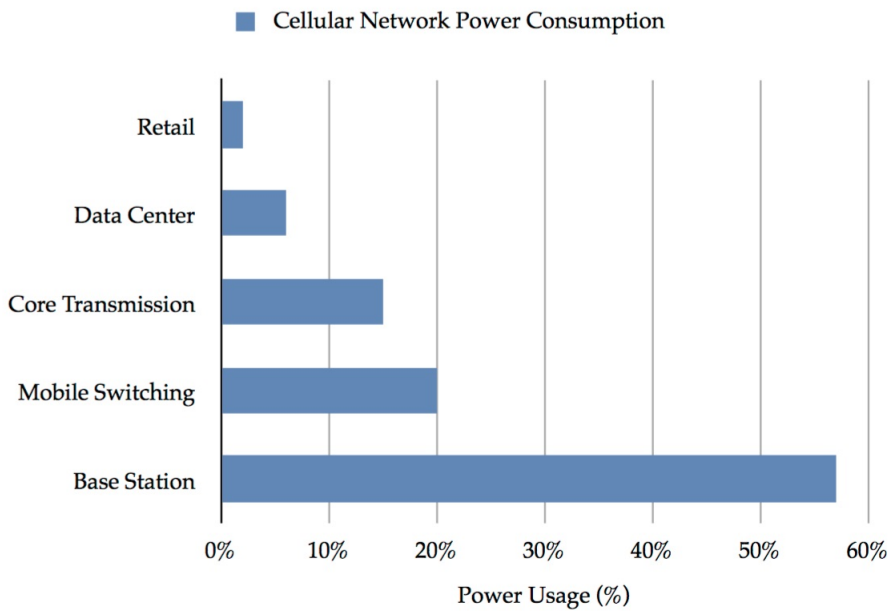


Figure 2.1: Cellular power consumption

digital-RF chains including the forward path and the feedback path, and a PA stage.

PAs are power gain units used to boost the signal power before it is transmitted over the air. This means the PA generates large amounts of power to allow for wider wireless signal coverage. The PA in a base station consumes the majority of the total energy(about 65%) [9], as the distribution of power consumption of radio base station shown in Fig. 2.3. Moreover, modern mobile radio systems which require radio transceivers able to support high data rates and throughput have strong linearity requirements. To achieve high efficiency for saving energy, the PA needs to work in the highly nonlinear peak power regime. On the other hand, in that regime, the PA exhibit nonlinear distortion that brings problems related to preserving high signal quality. The PA system comes with an inherent trade-off constraint between high efficiency and signal quality (PA linearity). Thus, as the dominant power consumer in a base station, the PA plays a crucial role in determining the overall power efficiency of a wireless transmitter and the

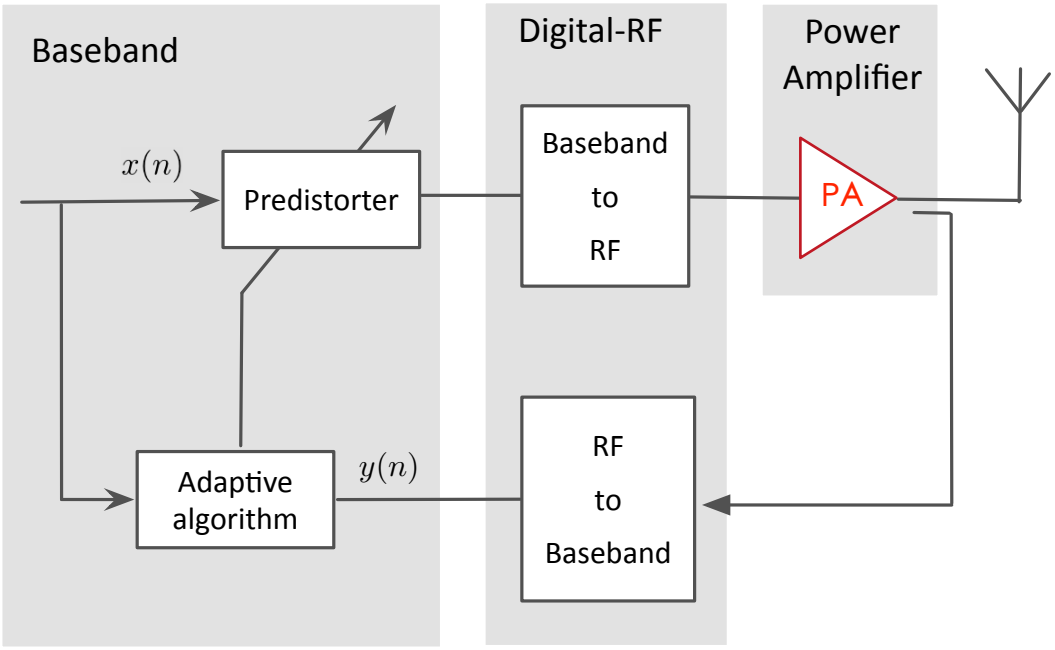


Figure 2.2: A simplified diagram of a high-efficient wireless transmitter.

quality of the transmitted signal.

The power efficiency can be measured by total efficiency or power added efficiency (PAE). The total efficiency gives a picture of the ratio of output power to both types of input power (DC and RF), written as

$$\eta = P_{out}/(P_{DC} + P_{in}) \tag{2.1}$$

where P_{in} and P_{out} indicate the input and output power of the PA, and P_{DC} indicates the DC power supply.

The PAE is a metric for rating the efficiency of a PA that takes into account the effect of the gain of the amplifier. PAE is often applied to amplifiers as a figure of merit, as well as devices.

$$PAE = (P_{out} - P_{in})/P_{DC} \tag{2.2}$$

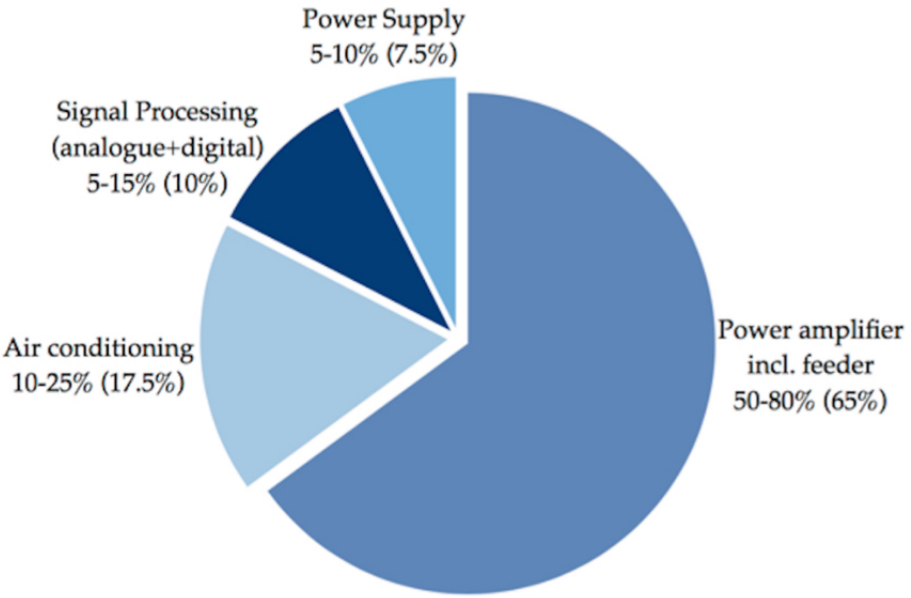


Figure 2.3: Base station power consumption

where P_{in} and P_{out} indicate the input and output power of the PA, and P_{DC} indicates the DC power supply. In a wideband macro-cell base station, a typical PAE of PAs maybe around 30%, but it is highly related to the PA architecture and PAPR and bandwidth of the transmitted signal. In [10], it shows that RF PAs can achieve some state of the art efficiencies from 49% to 70% PAE.

RF PA is the key component in wireless communication system, especially for the cellular base station. Linearity and efficiency are the two crucial design criteria because of signal quality and energy conservation. Specifically, the RF PA in a wireless transmitter should provide linear gain to meet any linearity specifications such as adjacent channel leakage ratio and error vector magnitude as defined by statutory bodies such as International Telecommunication Union (ITU). Furthermore, the RF PA has not only to be as efficient as possible, but also to have wide bandwidth to accommodate the signal bandwidth and pre-distortion signal bandwidth.

2.1.2 PA Classes

Not all amplifiers are the same. There is a clear distinction made between the way that their output stages are configured and operated. Amplifiers are classified into classes according to their construction and operating characteristics [11]. The class gives a broad indication of an amplifier's characteristics and performance. The classes A, B, AB and C refer to the way the amplifiers are biased, although class C is mainly used in the oscillator circuits. Classes D to H are used in switch mode amplifiers where power is saved by having the output transistors switched rapidly between fully on and fully off. In either of these states, the transistor is dissipating little or no power.

The purpose of the class A PA is to make the amplifier relatively free from distortion by keeping the signal waveform out of the region between 0V and about 0.6V where the transistor input characteristic is non-linear. In a Class A amplifier, the transistor is switched on for the whole of the signal cycle. By careful management of the signal levels, Class A amplifiers can be quite linear, but are wasteful of power. The output power they produce is theoretically 50%, but practically may be only about 25 to 30%. The bias of a Class A PA is shown in Fig. 2.4.

In terms of class B bias, there is no standing bias current (the quiescent current is zero) and therefore the transistor conducts for only half of each cycle of the signal waveform, shown in Fig.2.5. This dramatically increases efficiency, compared with class A. Theoretically nearly 80% efficiency can be achieved with this bias and in practical circuits, efficiencies of 50% to 60% are possible. The downside for this increased efficiency is that the transistor only amplifies half the waveform, therefore producing severe distortion.

The class AB push-pull output circuit is slightly less efficient than class B be-

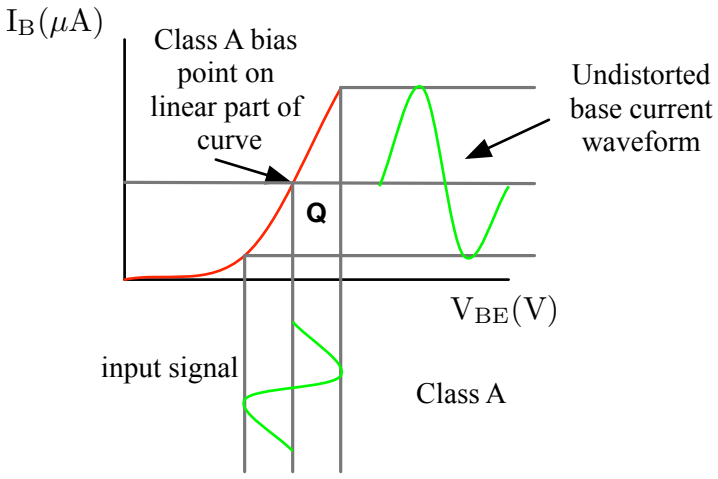


Figure 2.4: The operating bias and signal conduction of Class A amplifiers

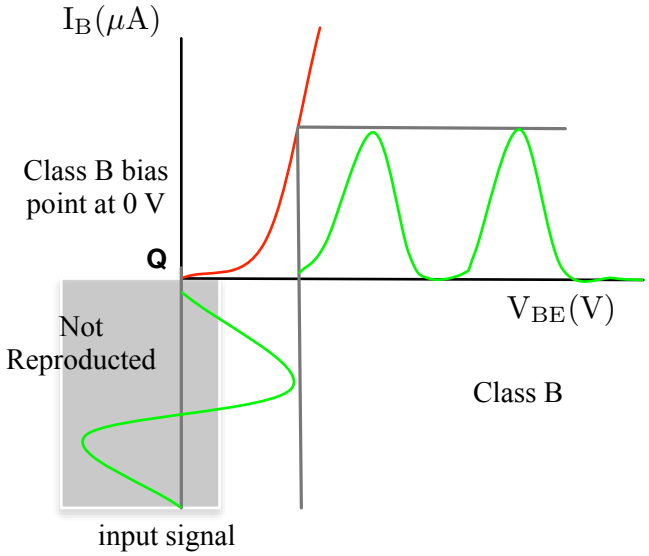


Figure 2.5: The operating bias and signal conduction of Class B amplifiers

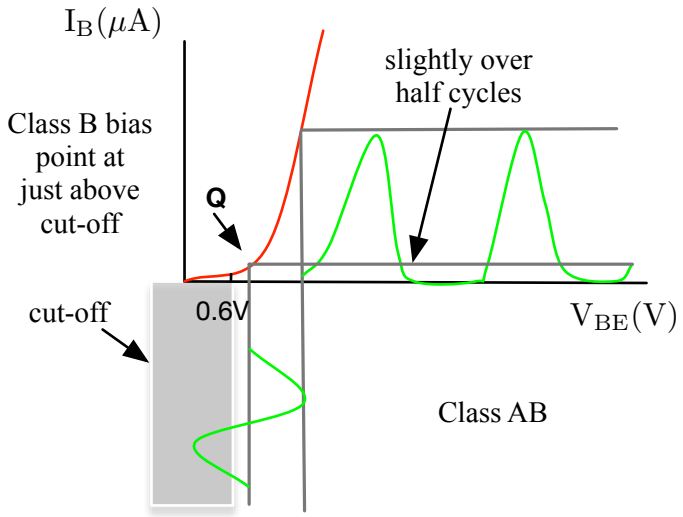


Figure 2.6: The operating bias and signal conduction of Class AB amplifiers

cause it uses a small quiescent current flowing to bias the transistors just above cut off, shown in Fig. 2.6. However, the crossover distortion created by the non-linear section of the transistors input characteristic curve which near to the cut-off point in class B is overcome. In class AB each of the push-pull transistors is conducting for slightly more than the half cycle of conduction in class B, but much less than the full cycle of conduction of class A.

In Class C, the bias point is placed well below cut-off so the transistor is cut-off for most of the cycle of the wave. This gives much improved efficiency to the amplifier, but very heavy distortion of the output signal.

Class D amplifiers work by generating a variable duty cycle square wave of which the low-frequency portion of the spectrum is essentially the wanted output signal, and of which the high-frequency portion serves no purpose other than to make the wave-form binary so it can be amplified by switching the power devices. Class D operation makes the output circuit extremely efficient (around 90%) allowing high power output without the need for high power transistors and elaborate heat-sinks. However, this big increase in efficiency is only achieved at the expense of some increase in distortion and especially of noise.

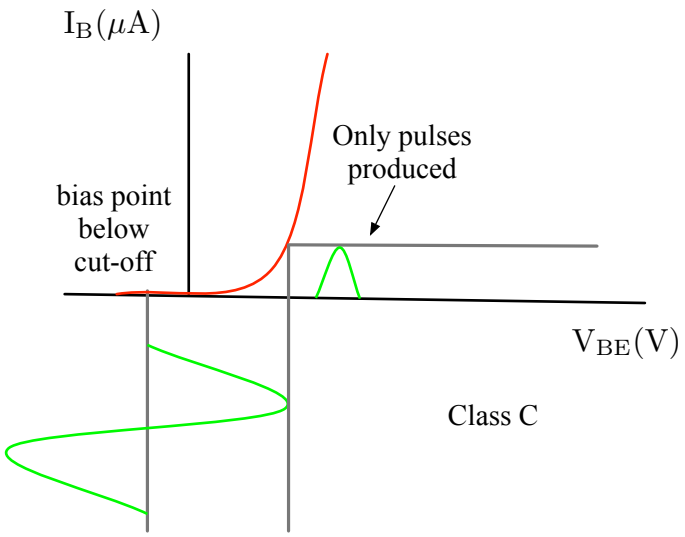


Figure 2.7: The operating bias and signal conduction of Class C amplifiers

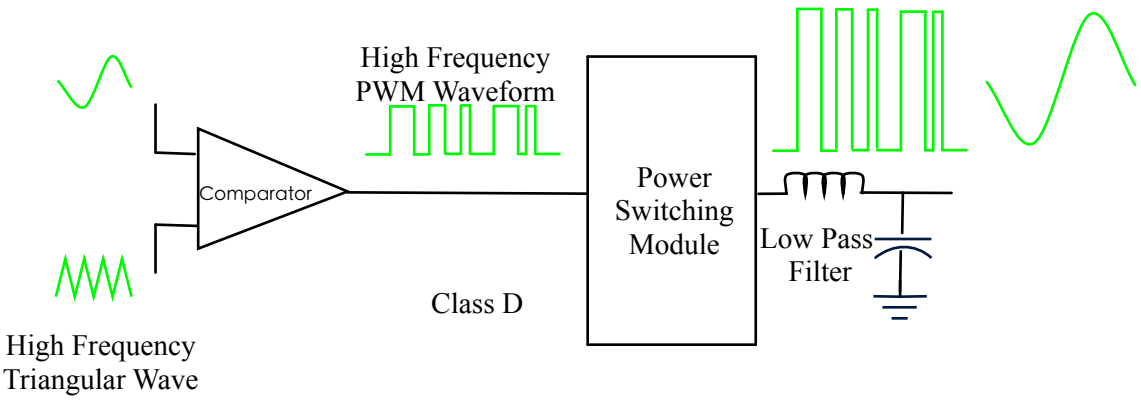


Figure 2.8: The operating bias and signal conduction of Class D amplifiers

2.1.3 High Efficient PA Architecture: Envelope Tracking and Doherty

In order to solve the dilemma on trade-off between efficiency and linearity in RF PA performance of base stations, there are two main classes of technique for improving RF PA efficiency for high-PAPR signals: drain modulation and load modulation. Drain Modulation refers to a set of techniques in which the drain supply voltage of the amplifier is adjusted to follow the envelope of the signal, as shown in Fig. 2.9. The drain supply is set dynamically so that the amplifier operates close to power saturation at all times. Envelope Tracking (ET) is a one of the drain modulation technique that requires the supply voltage of the final PA stage to be modulated dynamically with the envelope of the input signal [12]. Traditional constant PA supply voltage has significant amount of power loss as heat. Fig. 2.10 presents that envelope tracking techniques modulate the supply Voltage to track the envelope of the signal to reduce the amount of power dissipated as heat. This would make the PA operate closer to the peak level at all times and dramatically improve the efficiency of the PA. Overall System efficiency using ET techniques must include the drain modulator. In an ET system, modulator efficiency normally is around 80%, and the amplifier efficiency must be more than 50%. The efficiency of a ET system should be at least 10% better than classical PA classes. Recently ET techniques attract more attention in handset applications. Challenges of ET PAs are to increase the output power and signal bandwidth [13].

Doherty PA was developed in 1936 [14], and it increases the average efficiency of the linear mode PA by parallel combination of their output powers as shown in Fig. 2.11. The Doherty amplifier consists of two amplifiers operates at 90 degrees phase shift on both the input and output of its topology. Fig. 2.12,

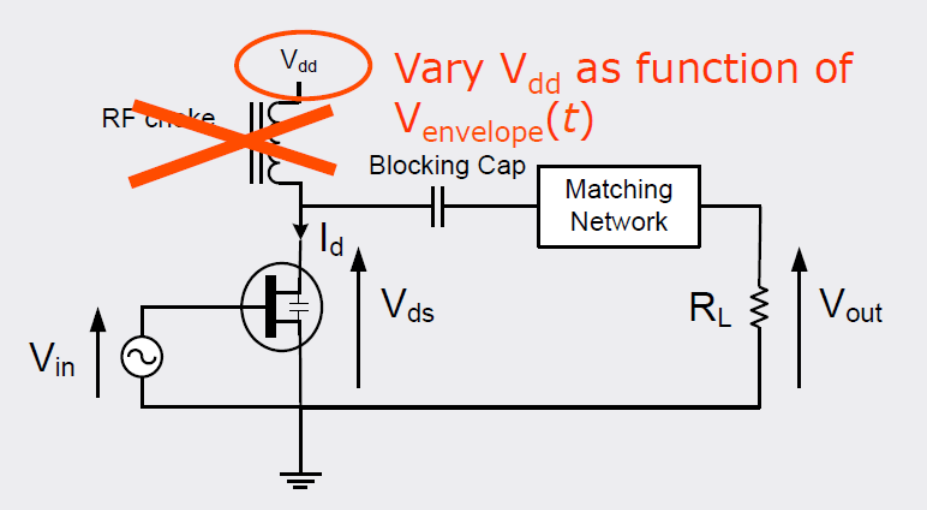


Figure 2.9: Diagram of the drain modulation amplifier

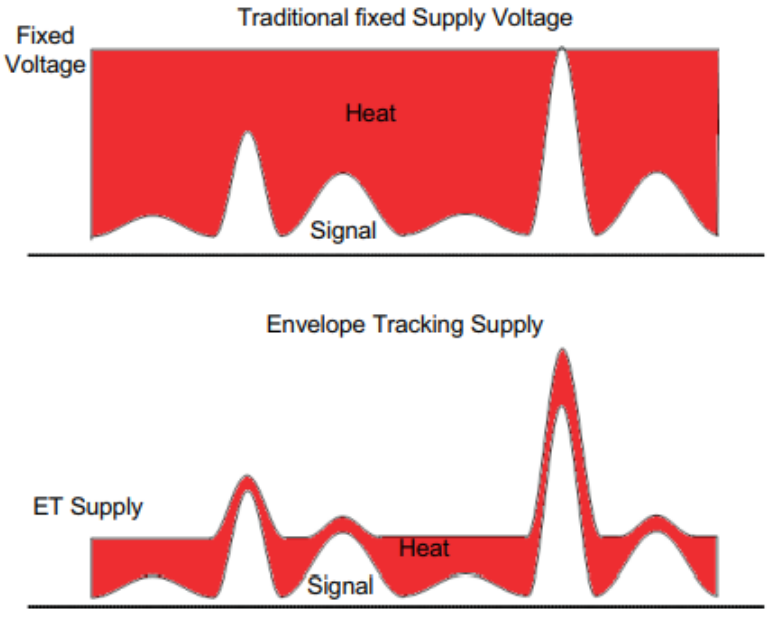


Figure 2.10: Traditional Supply vs Envelope Tracking Supply in [1]

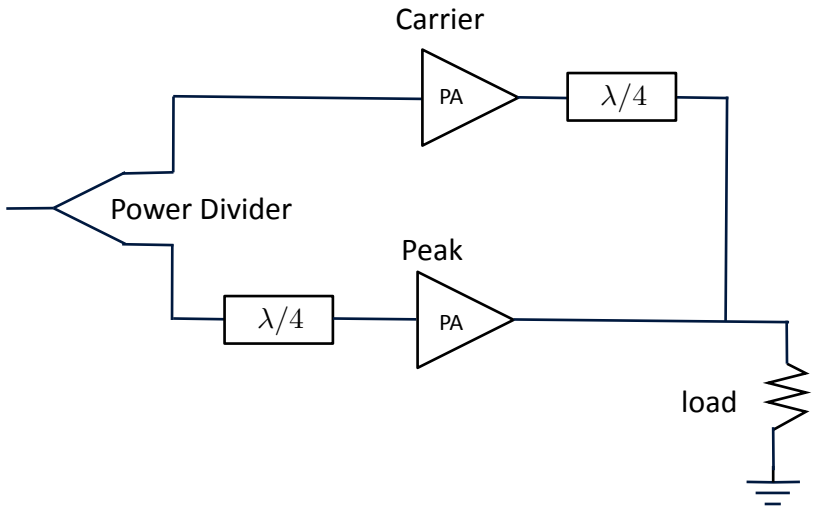


Figure 2.11: Doherty transmitter

introduced in [2], depicts the measured gain and power-added efficiency (PAE) of the balanced class-AB and Doherty amplifiers with a continuous wave (CW) signal. It shows that the Doherty PA has higher efficiency at backed-off power levels than a balanced class-AB PA. The Doherty PA is currently the PA architecture of choice for cellular (Macro) base stations, DVB-T transmitters and many applications using high-PAPR signals. The challenges of Doherty PAs are the signal bandwidth and further improving the efficiency. For a Doherty PA, it is difficult to increase the signal bandwidth greater than 20% of central frequency. The multi-band Doherty PA which covers several LTE bands with one PA is of interest for modern base stations.

2.1.4 Nonlinearity and Memory-effect

Since the efficiency concern of RF PAs attracts a lot of attention in recent decades, the aspect of efficiency becomes more and more critical with regard to the market requirement and environmental protection policies. In order to achieve high efficiency, the operating points of PAs must be pushed higher into

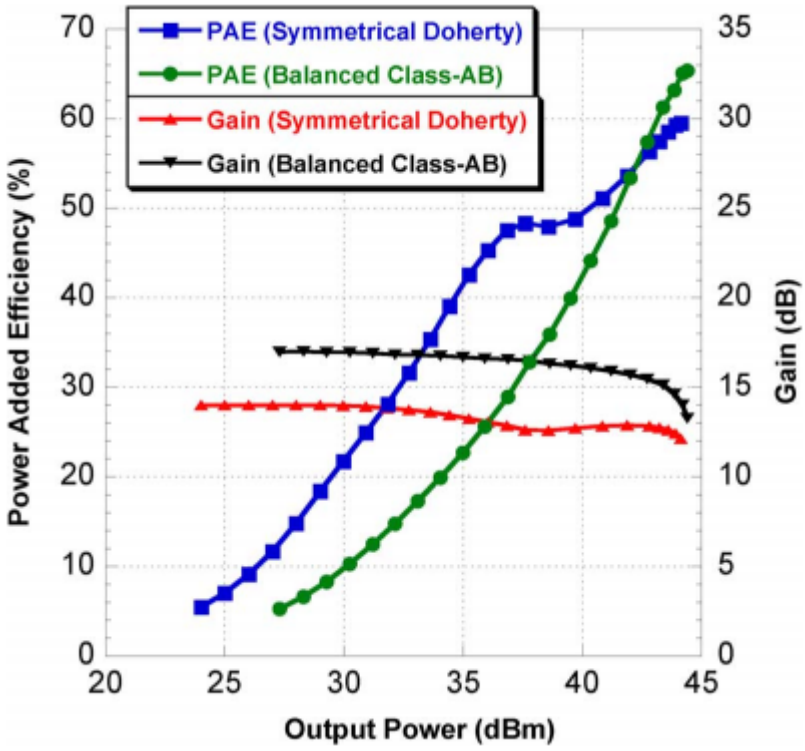


Figure 2.12: Measured PAE and gain of the balanced class-AB PA and symmetrical Doherty PA using a CW signal in [2]

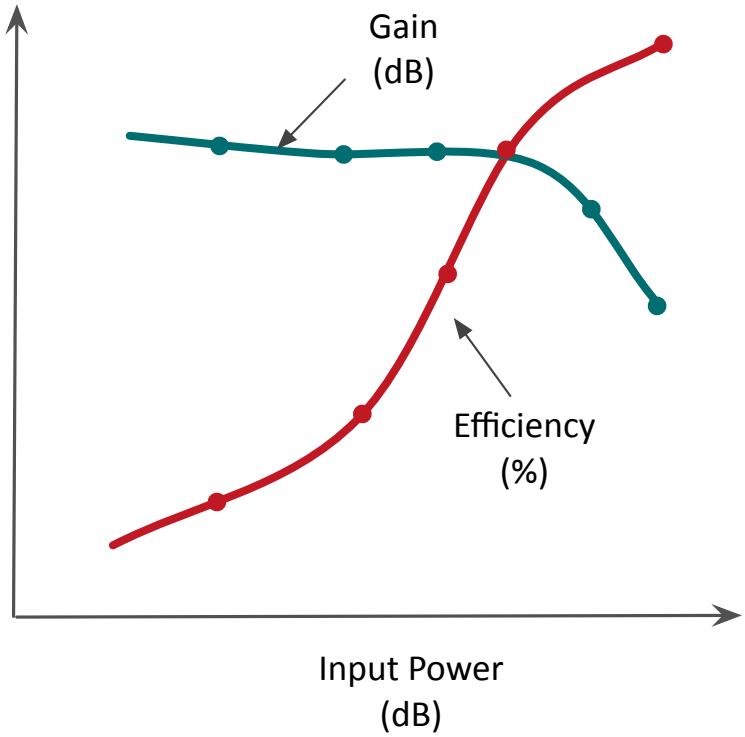


Figure 2.13: Typical PA efficiency gain characteristics

the saturation (nonlinear) region. This means that the PA achieves typically a higher efficiency at a higher input and output power level. However, the gain will be compressed with the large input power levels that drive the PA into the nonlinear region. The gain and total efficiency characteristics of a PA are presented in Fig. 2.13. The compressed gain causes the nonlinearity of the system and the distortion of the output signal. Thus, the high efficient PA is identified as the major source of nonlinear distortions in a wireless transmitter. The design of PA systems is always subject to the unavoidable antagonism between linearity and power efficiency.

An amplifier sustains a constant gain with low-level input signals. With higher input power, the PA goes into saturation and its gain decreases. The 1 dB compression point (P1dB) denotes the power level that causes the gain to drop by 1 dB from its small signal value. IP3 is another well-known metric that

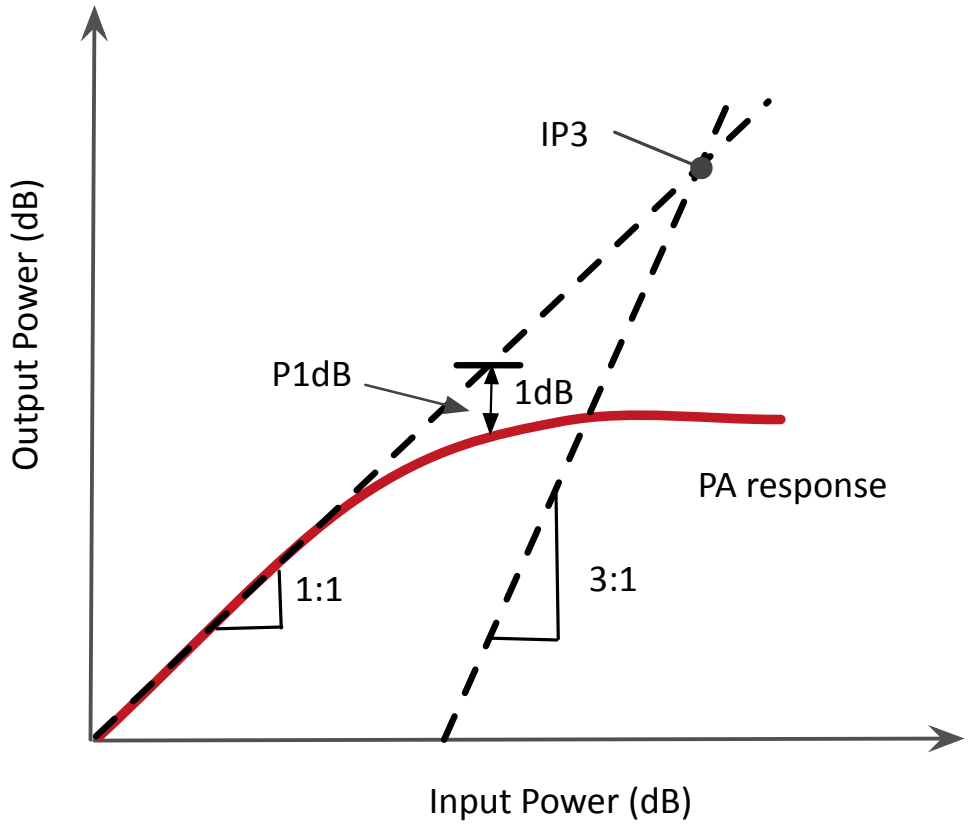


Figure 2.14: P1dB and third order intercepts points

gauges linearity of the PA. A higher IP3 indicates better linearity and lower distortion. It is the theoretical point at which the desired output power is the same as the power of undesired third-order intermodulation signal. The P1dB, IP3 point and input power vs output power characteristic can be read off from Fig. 2.14.

Memory effects are the other source that cause distortions, besides the gain compression. Memory effects can be simply explained by the fact that the output of the system exhibits latency. The output signal at this instant is not only determined by the input signal at this instant, but also by the input signal at the previous instants. The normalized output amplitude against the normalized input amplitude is defined as Amplitude-to-Amplitude (AM-AM)

characteristic, which displays the output amplitude distortions. In addition, the phase distortion is described by Amplitude-to-Phase (AM-PM) characteristic, where the phase difference between input and output is shown in the plot with regard to the normalized input amplitude. The Memory effect leads to a non-flatness in the PA output spectrum, and a spreading of points at AM-AM and AM-PM curve, as shown in Fig. 2.15. Briefly, memory effects can be classified into short term and long term memory effects according to the involved time constants of the memory effects. The short-term memory effects have time constants about several times of carrier signal period. In contrast, long-term memory effects exhibit much lower frequency, typically from a few kHz to MHz [15].

The nonlinearity and memory-effects of the PA cause the appearance of spectrum regrowth at the output of the transmitter, which is also called intermodulation distortion(IMD). The frequency components of IMD present at the output of the nonlinear PA driven by a two-tone input signal are summarized in Table 2.1, and illuminated in Fig 2.16. The DC components, the second and third order harmonics, second-order inter-modulation distortions, as well as out-of-band third order inter-modulation distortions can be easily filtered out by a band-pass filter around the fundamental frequency. Therefore, their significance is minor. On the other hand, the distortions appearing at the same frequencies as the input signal, and the in-band third-order IMD products that are too close to fundamental components to be filtered. In practical, if the continuous wideband signal is used, IMD components appear as a spectrum regrowth around the signal, as shown in Fig. 2.17. The system interference budgeting requires transmitted IMD levels in the adjacent and alternate channels be limited so as not to cause interference to other radios operating in nearby channels.

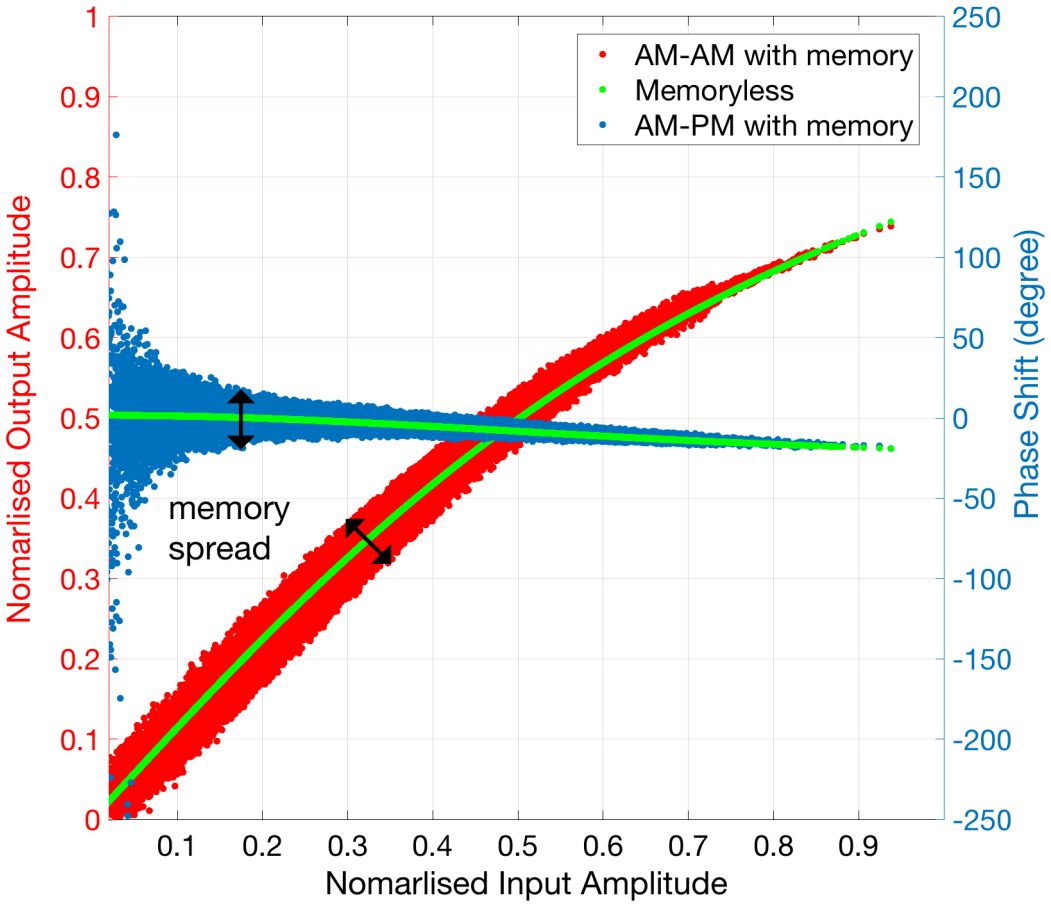


Figure 2.15: Example of an AM/AM and AM/PM curve for a PA with memory

Table 2.1: Frequency components of some harmonics IMD with a two-tone input

Frequency Locations	Designation
0	DC components
ω_1 and ω_2	Fundamental
$2\omega_1$ and $2\omega_2$	Second harmonics
$3\omega_1$ and $3\omega_2$	Third harmonics
$\omega_2 - \omega_1$ and $\omega_2 + \omega_1$	2^{nd} IMD
$2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$	In-band 3^{rd} IMD
$2\omega_1 + \omega_2$ and $2\omega_2 + \omega_1$	Out-band 3^{rd} IMD

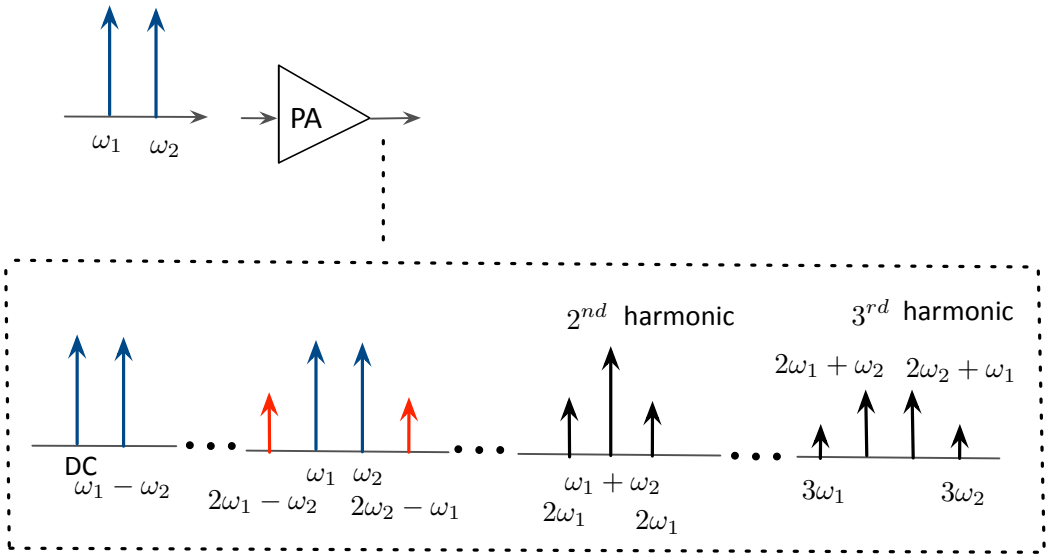


Figure 2.16: Frequency domain output of a nonlinear PA driven by a continuous wideband signal

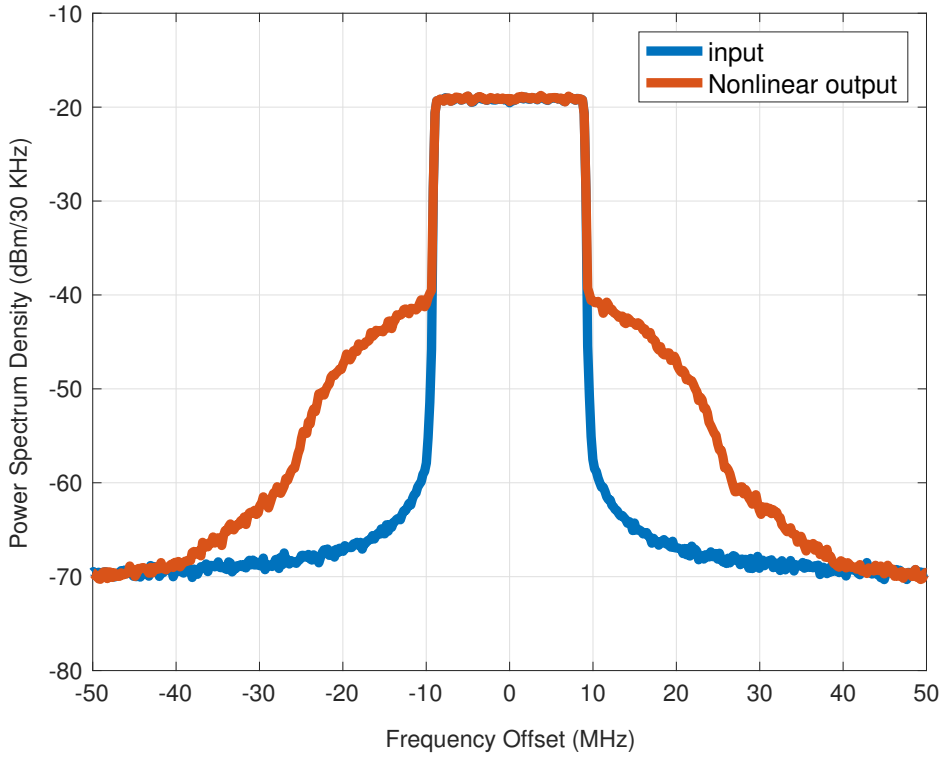


Figure 2.17: Frequency domain output of a nonlinear PA driven by a two-tone signal

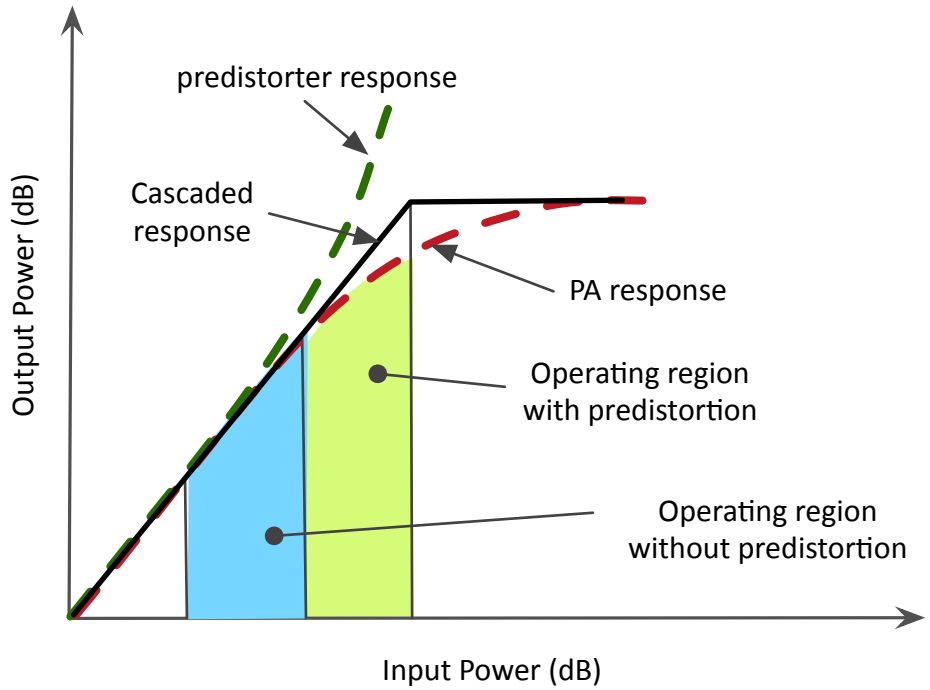


Figure 2.18: The DPD operation principle

2.2 Digital Predistortion Techniques

Driving PAs to the nearly-saturated region increases the power conversion efficiency while leading to the spectrum regrowth due to the nonlinearities correspondingly. The trade-off between linearity and power efficiency of the PA is a challenging topic in the system design. During the past decade, DPD techniques have proven to be one of the most effective solutions to achieve the adequate linearity of RF PAs [7]. By reducing the distortion created by operating PAs in their non-linear regions, PA can be made to be far more efficient. The principle of DPD is to intentionally introduce a nonlinear function in the baseband processing, which is commonly called the predistorter path. Ideally, the cascade of the predistorter and the PA becomes linear and the original input is amplified by a constant gain as shown in Fig. 2.18.

To design a DPD system, a proper behavioral model must be defined first.

Behavioral modeling uses a series of mathematical formulations to relate input and output signals. It provides a convenient and efficient means to predict system-level performance without the computational complexity of full circuit simulation or physics-level analysis of nonlinear systems, thereby significantly speeding up the analysis process [16]. Fig. 2.19 shows a DPD-enabled wireless transmitter, where the DPD system is in the shaded boxes. A DPD system, in general, is comprised of three main blocks: predistorter path, coefficients training and an observation receiver (also known as the feedback path which is used to capture the PA output signal for calculating DPD coefficients). The predistorter, where we use the behavioural model to represent the nonlinear PA or the nonlinear transmitter, is a part of digital baseband processor system. It generates the output of the behavioural model as the predistorted signal. It is clear that the behavioral model is crucial to predict the nonlinearity of the PA in the transmitter in general. The key advantage of behavioral modeling resides in the fact that it does not require deep knowledge of the RF circuit physics and functionality. The synthesis of the predistortion function is equivalent to the behavioral modeling of the PAs reverse function obtained by swapping the PAs input and output signals with appropriate small-signal gain normalization [17].

The sample rates in baseband can be several hundred MHz according to the signal bandwidth and spectrum expansion. Then the digital predistorted signal is converted to the analog signal by the digital-to-analog converter (DAC), and up-converted to the RF frequencies by the IQ-modulator or mixer. To calculate the coefficients of the behavioural model, the output samples of PA is required. The attenuated output signal is first down-converted to the IF or baseband frequencies, then captured by the analog-to-digital converter (ADC). These analog paths must be able to cover the wider signal bandwidth of the predistorted signal, due to the added distortion components. The coefficients of

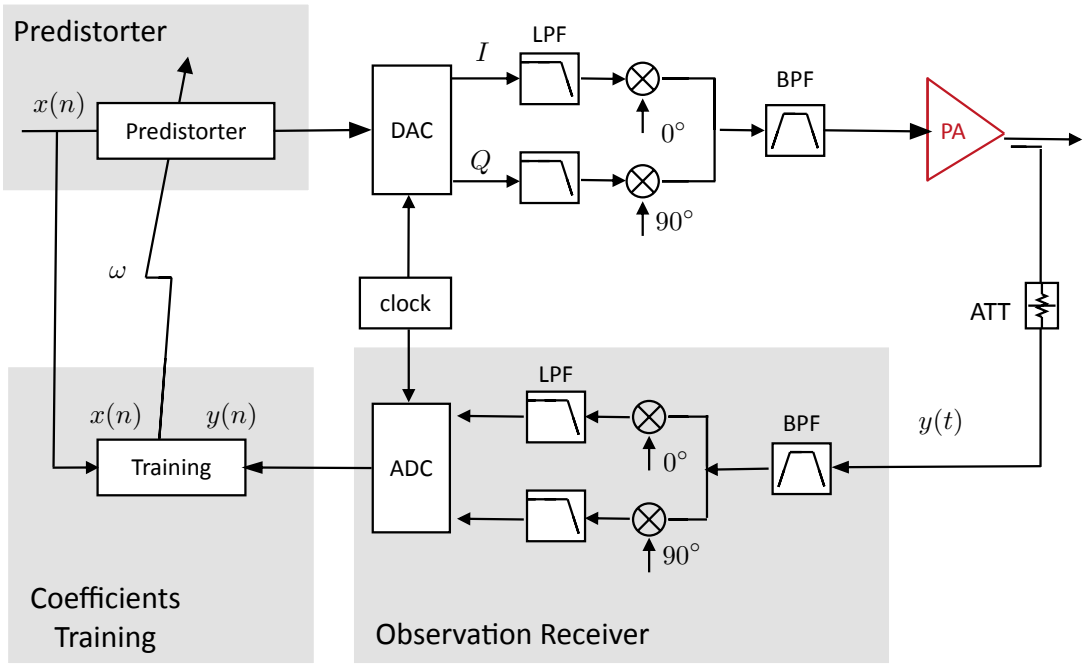


Figure 2.19: A simplified diagram of a DPD-enabled wireless transmitter.

the behavioural model are typically obtained using least-square (LS) based algorithm. The accuracy of the DPD system is briefly determined by the behavioural model, digital-RF paths and the coefficient extraction.

2.2.1 Behavioural Models

Volterra Model

Volterra Model is the most comprehensive model for dynamic nonlinear system [18]. It is a polynomial approximation to the target function at the present instant convolved with the target function over some history. Volterra model can also be treated as a Taylor series with the memory.

$$y(n) = \sum_{k=1}^{K} \sum_{i_1=0}^{M} \cdots \sum_{i_p=0}^{M} h_p(i_1, i_2, \cdots, i_p) \prod_{j=1}^{k} x(n - i_j) \tag{2.3}$$

where $h_p(i_1, i_2, \cdots, i_p)$ are the coefficients of the Volterra model, K is the nonlinearity order of the model, and M is the memory depth.

The Volterra based models have demonstrated high accuracy in modelling nonlinear PAs and transmitters [17]. One major drawback of the classical Volterra series is its large number of coefficients. The number of coefficients increases exponentially with the nonlinearity degree and the memory. It limits the practical use of the classical Volterra series in the DPD system.

Wiener Model and Hammerstein Model

The short-term memory effects are caused mainly by the reactive components of the active devices and matching networks [15]. One phenomenon of memory effects is that the IMD components generated by PAs do not remain constant when varying the tone spacing with a Two-Tone test signal. Because of the frequency-dependent PA behaviour, one can directly think of uneven frequency response of input matching network of PAs. The input matching network can be represented by a linear time-invariant (LTI) system. As a consequence, this memory PA model can be represented by a cascade of an LTI system and a memoryless nonlinearity (NL), which is the so-called Wiener model [18], [19], as shown in Fig. 2.20.

The wiener model with two subsystems are given by:

$$u(n) = \sum_{l=0}^{L-1} a_l z(n-l), \tag{2.4}$$

$$y(n) = \sum_{k=1}^{K} b_k u(n)|u(n)|^{k-1} \tag{2.5}$$

Similarly, the PAs have an output matching network as well, which has certain

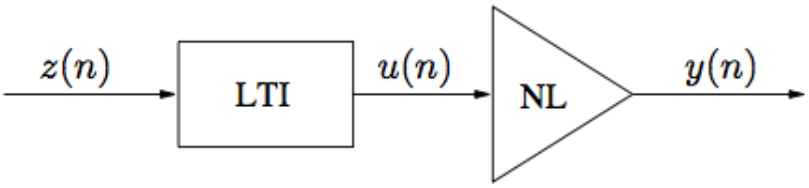


Figure 2.20: The wiener model.

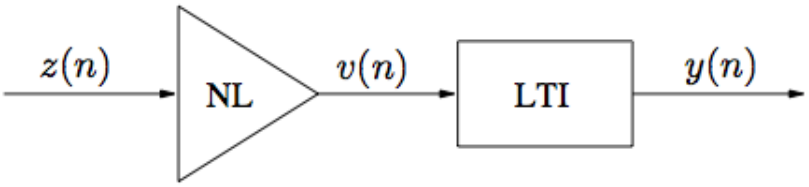


Figure 2.21: The Hammerstein model.

bandwidth limitation and ripples in its frequency response. A memoryless PA model followed by an LTI system has the name of Hammerstein model [20], as shown in Fig. 2.21. The Hammerstein model is written as,

$$\nu(n) = \sum_{k=1}^{K} a_k z(n)|z(n)|^{k-1}, \tag{2.6}$$

$$y(n) = \sum_{l=0}^{L-1} b_l \nu(n-l), \tag{2.7}$$

Generalized Memory Polynomial

One of the most popular simplifications of the Volterra model is the generalized memory polynomial (GMP) [21].

$$y(n) = \sum_{k\in K}\sum_{p\in P}\sum_{q\in Q} \omega_{kpq} x(n-p)|\ x(n-q)\ |^{k} \tag{2.8}$$

It divides memory-effects into base memory and cross memory. It is the base

memory term when p equals to q. Otherwise, it is the cross memory term. The base-term utilizes the diagonal kernels of the Volterra series corresponding to the pure powers of the input signals samples. The main advantage of this cross-term model is that the coefficients, like those of the memory polynomial, appear in the linear form. The GMP model is known to have excellent performance regarding linearity versus complexity [22].

2.2.2 Model Estimation

DPD linearization performance is not only depending on the accuracy of the DPD model, but also depending on the accurate of the model identification. The nonlinearity exhibited by the nonlinear system varies with time due to changes of the drive signal, aging, or drifts, thus it is essential to continuously update the behavioural model to maintain the linear operation of the system made of the predistorter and the PAs.

Generally, the DPD structure could be categorized into two separate classes, that is, the indirect learning structure (IDLA) [23] and the direct learning structure (DLA) [24], as shown in Fig. 2.22. The IDLA is based on the theory that the post-inverse behavior can be equally treated as the pre-inverse behavior [18]. The IDLA technique is known as fast converge after only two or three iterations which can Direct learning means that the model of the input-output relation of the PA is estimated first, and the predistortion is obtained directly by pre-inverting this PA characteristic. It extracts the coefficients by directly minimizing the residual error between the PA output and the original input. The DLA, in general, has better performance than the IDLA which is more sensitive to the feedback noise because the nonlinear matrix is based on the noisy feedback signalf [25] [26]. However, the accuracy improvement is normally at the price of a greater number

of iterations to converge.

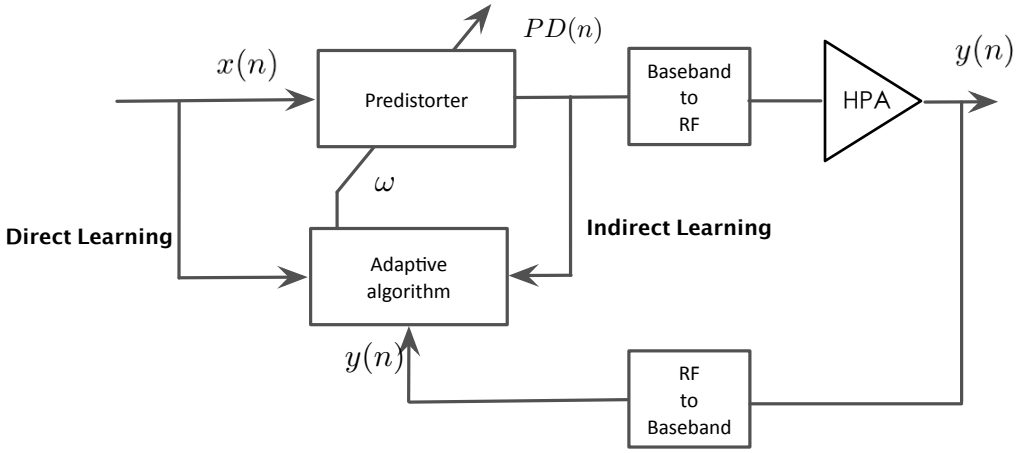


Figure 2.22: Direct learning architecture and indirect learning architecture for coefficients estimation.

Fig. 2.23 indicates a simplified estimation procedure. To estimate the accuracy of the coefficients of DPD, because of the fixed-point calculation and the measurement error in practice, a large number of training samples is required to implement the practical DPD estimation, which is normally from 8 thousand to 32 thousand samples. Then the alignment operations (or calibration) for correcting the time, phase and gain mismatch are essential to obtain the aligned training samples set. Finally the least-square (LS) based algorithms, e.g. LS, recursive least square (RLS) and least mean square (LMS) are normally employed to extract the coefficients using training samples.

2.2.3 Linearization Performance Measurement

The most straightforward metric for model performance assessment is to evaluate the discrepancy between the desired original signals and estimated output signals in the time domain. The normalized mean square error (NMSE) and error vector magnitude (EVM) have been regularly utilized for the performance assessment of

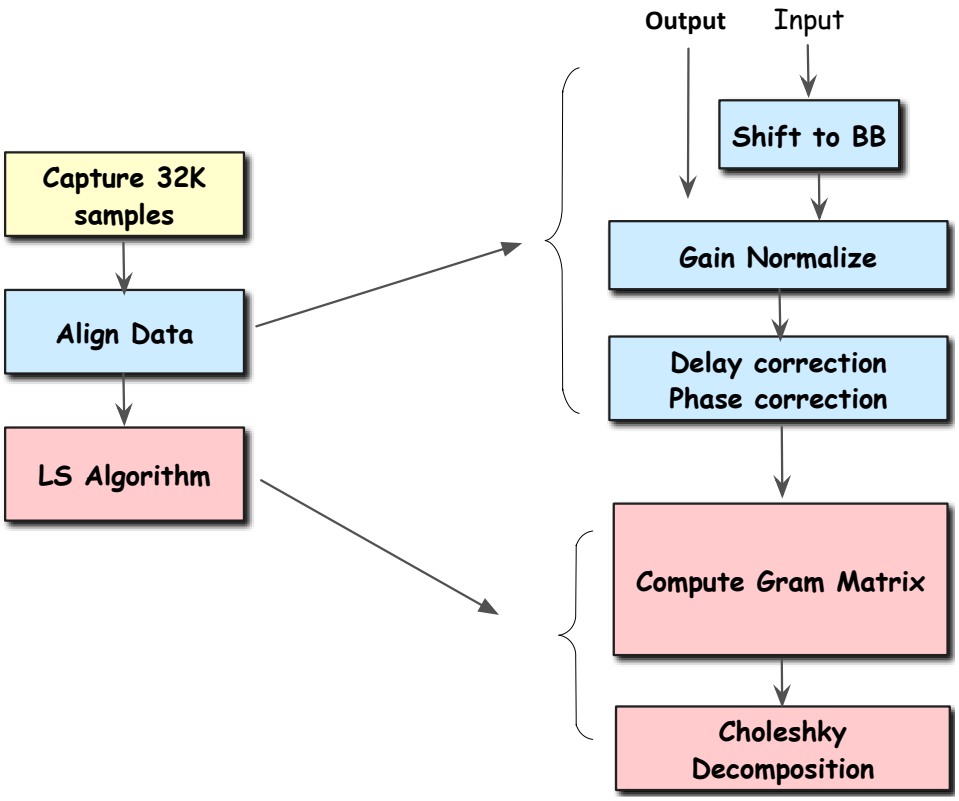


Figure 2.23: A simplified diagram of a model estimation procedure of DPD

behavioral models and DPD, which effectively evaluate the in-band nonlinearity. EVM is measured for the demodulated I/Q signal; NMSE is measured for the raw time domain signal before demodulation, shown in (2.10).

$$EVM = \sqrt{\frac{1}{N}\frac{\sum_{n=1}^{N}||y(n)-x(n)||^2}{||x_{max}||^2}} \tag{2.9}$$

where x_{max} is the sample corresponding to the highest amplitude point among the ideal input signal, N is the total number of samples.

$$NMSE = 10log_{10}\left(\frac{\sum_{n=1}^{N}|y(n)-x(n)|^2}{\sum_{n=1}^{N}|x(n)|^2}\right) \tag{2.10}$$

where N is the total number of samples

Out-of-band distortion is measured by the power of distortion falling into adja-

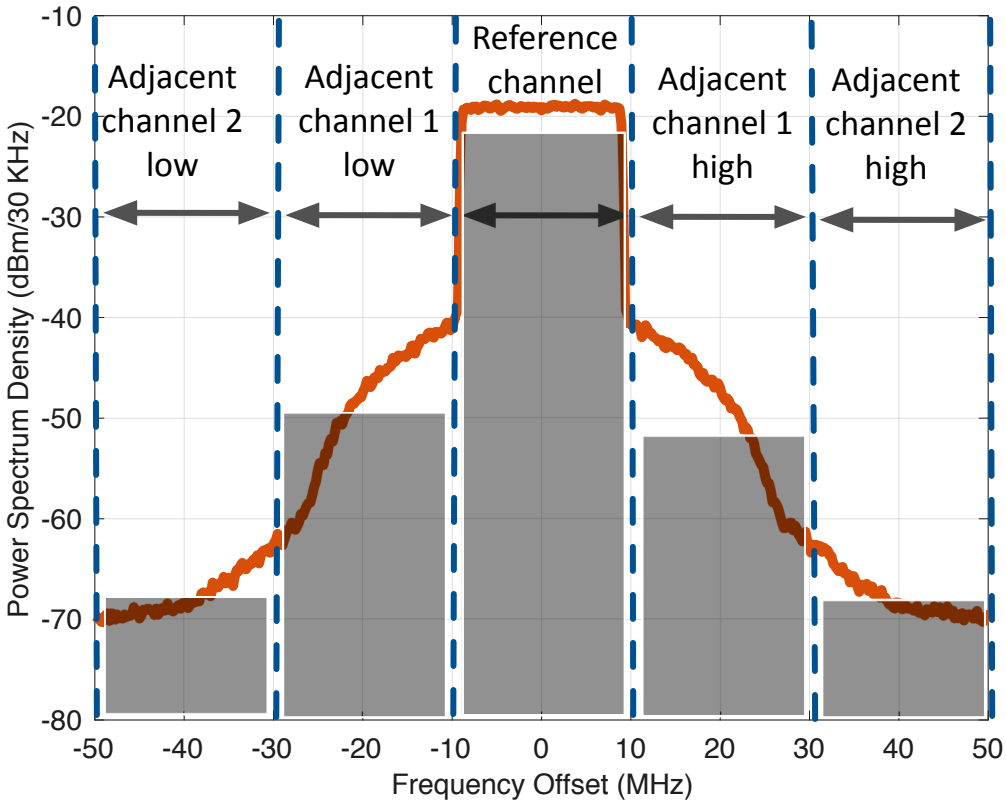


Figure 2.24: A diagram of ACPR measurement

cent channels near the allocated channel. This can be described as the adjacent channel power or the adjacent channel leakage. The ratio of the adjacent and allocated channel powers is referred to as the adjacent channel power ratio (ACPR) or the adjacent channel leakage ratio (ACLR), which can be used to effectively assess the out-of-band distortion shown in Fig.2.24.

2.3 Wideband and Multi-band Digital Predistortion Review

The complexity of DPD systems is increasing and becoming more challenging as the signal bandwidth and the number of fragmentary bands increase. This

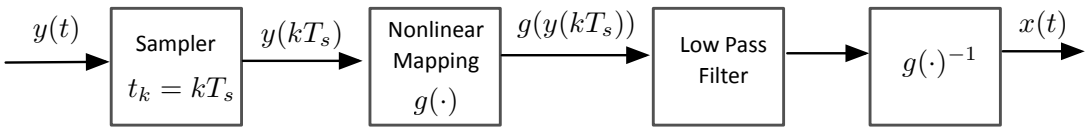


Figure 2.25: Diagram of the generalized sampling theorem.

section reviews DPD techniques in the area of wideband and multi-band applications.

2.3.1 Wideband Digital Predistortion

To address the challenge of the requirement of high-speed ADCs in the DPD system, considerable effort has been made [27–41].

In 1992, Yangming Zhu proposed the generalized sampling theorem (ZGST) that it is possible to employ undersampling in the modelling and identification of a band-limited nonlinear function [27]. ZGST intends to reconstruct undersampled signals with the aid of a function that reduces the bandwidth as well as its inverse, shown in equation (2.11) and Fig. 2.25.

$$f(t) = \sum_{k=-\infty}^{+\infty} g(f(t_k))sin[\omega_0(t - kT_s)]/[\omega_0(t - kT_s)] \tag{2.11}$$

As shown in Fig. 2.26, the sampling frequency requirement for the nonlinear system identification and compensation can be determined by the application of Zhu's generalized sampling theorem. Subsequently, some researchers study this theorem under the Volterra system specifically. For the identification of Volterra systems, in 1996, [42] indicated that it is sufficient to sample the continuous-time input and output signals at twice the maximum input frequency. Furthermore, Tsimbinos in 1998 proved that all of the nonlinear systems can be identified and compensated at the lowest possible sampling rate: the Nyquist rate of the input

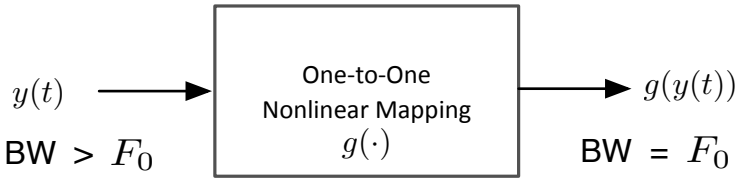


Figure 2.26: Bandwidth reduction by nonlinear mapping

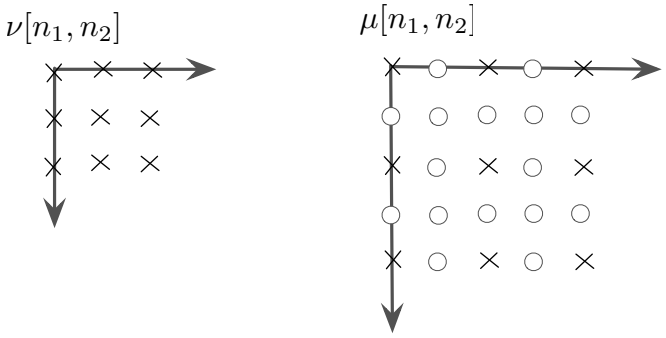


Figure 2.27: Kernel interpolation method

signal [28].

Using Zhu's under-sampling theorem, Peter Singerl and Heinz Koeppl first implemented the low-rate identification to the DPD system [29] [30]. It estimates the coefficients under a low-rate model, and employs a Volterra kernel interpolation method to recover the low-rate coefficients to high-rate coefficients for the high-rate predistorter, shown in 2.27. In 2008, Anding Zhu employed a similar method [31] for the DPD application. However, these approaches are only suitable for weakly nonlinear systems because it down-samples the input signal to the low-sampling-rate signal for coefficient estimation. This leads to the loss of information of memory variables in behavioural models.

To avoid down-sampling the input signal, [32–34, 36] proposed low-rate DPD techniques by compensating or recovering captured low-rate feedback samples to high-rate samples through additional modeling approaches. Then the coefficients are estimated from the high-rate input and high-rate feedback samples as the same as traditional approaches. As the spectra outside the acquisition

band are filtered out by the narrow band-limiting filters, the demands on ADC speed can be much lower for obtaining the un-aliased signal. Braithwaite [43] explored a DPD architecture by using multiple narrow-bandwidth observations to reduce the feedback bandwidth. Liu [33] proposed an under-sampling restoration method to restore the full-band feedback signal from a narrow-band feedback path by a dual-loop iteration procedure. It restores the full-band signal using a time-domain method, but it requires an additional DAC for training. Ma [34] proposed a spectral extrapolation method to recover the band-limited signal in frequency-domain. It is the first paper that demonstrates the bandwidth of feedback signal can be even lower than the bandwidth of the original signal. Hammi [44] proposed a two-box DPD architecture based on the cascade of a memory polynomial (MP) and a memoryless predistortion function. By modeling the memoryless static nonlinearity offline from narrow-band measurements the observation path is only used to identify the memory-effect information which leads to the reduction of the sampling rate. Teng [45] proposed a multi-rate method to calculate the coefficients with a high-rate and low-rate training sample in direct learning architecture. Zonghao [36] proposed a multi-rate identification method to recover the feedback signal by an indirect learning architecture.

The ADC sampling rate determines the maximum pre-distortion signal bandwidth that can be accommodated. Sub-sampling approaches can be used to relief the need for the high-speed high-resolution ADC in the observation receiver. The DPD coefficients identification can be carried out by comparing the input and captured signals at a low sampling rate. The ZGST principle used in the application of PA behavioral modeling requires a priori knowledge of its corresponding inverse function. In this way, the PA output can be captured using the input Nyquist rate (twice the input bandwidth), rather than output

Nyquist rate. The applications [46] [47] [29] [30] [31] applying ZGST for the DPD identification perform well in some low-PAs which the memory-effects are considerable short and weak. Various advanced approaches are proposed to estimate high power PA models or inversed high power PA models using low-rate feedback signals. However, most of approaches are based on different principles and some of them requires considerable additional calculation or hardware resources. Some recent work shows that the low-rate DPD based on the aliasing effect of undersampling PA output is more promising in the theory and experimental results [38] [40] [41] [36]. A undersampling factor of 20 is reported [48] [36] [41]. However, in most of the low-rate DPD work, ADCs are operated in a high sampling-rate, such as hundreds of million samples per second(MSPS). Then low-rate DPDs are performed using the digital down-sampled feedback signal to mimic the captured signal from low-rate ADC . More recently, direct learning DPD systems have been investigated in which theoretically the RF-PA feedback signal can be sampled at an arbitrary low-rate [38]. However, even though the feedback samples can be arbitrary low-rate, it doesn't indicate that the low-rate ADC can be directly employed in DPD systems. The analog bandwidth of ADC is still needed to be greater than the Nyquist frequency of PA output signal which is also reported in [38]. This requirement limits the choice of ADC, since ADCs with large analog bandwidth are also tend to be high-speed. A recent solution [41] to alleviate this bottleneck is to insert a multi-rate track-and- hold amplifier (THA) before ADC.

2.3.2 Multi-band Digital Predistortion

Multi-band and multi-standard wireless communication systems have caught the attention of recent research activities. It is potentially a cost-effective solution for higher data rates and more optimal use of the frequency spectrum. In multi-

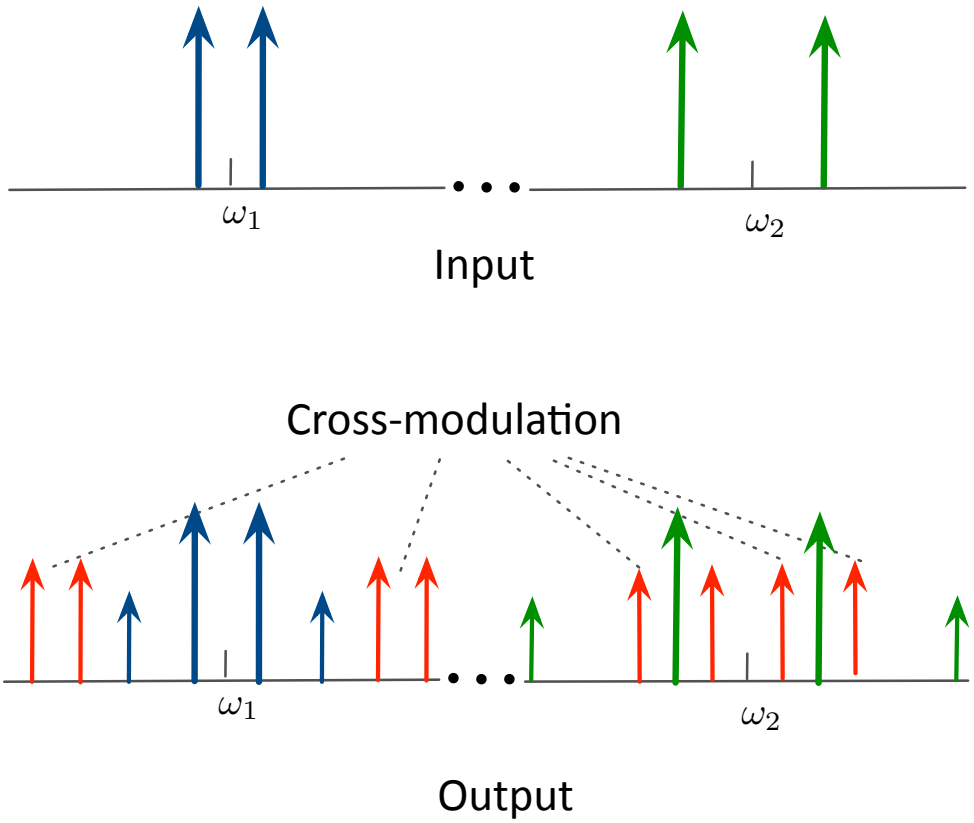


Figure 2.28: Effect of PA nonlinearity on concurrent dual-band transmission

band transmitters, the RF front-end hardware for the concurrent transmission of various wireless signals at different carrier frequencies is the same. The concurrent multi-band transmission exhibits the distortion challenge in the nonlinear components of the transmitter, mainly the PA. Besides the inter-modulation products, cross-modulation products are generated by multi-band signals, as shown in Fig. 2.28.

There are several challenges for the design of the concurrent multi-band transmitting systems. One of the main issues is that the separation of the carrier frequencies can be very large. For an instance, Fig. 2.29 shows the noncontiguous signals in band 1 and band 3 with a mix of LTE and UMTS modulations. With DPD bandwidth expansion, we are looking at 500 - 600 MHz of total bandwidth. It requires the impractical sampling rate requirement for the analog-to-digital

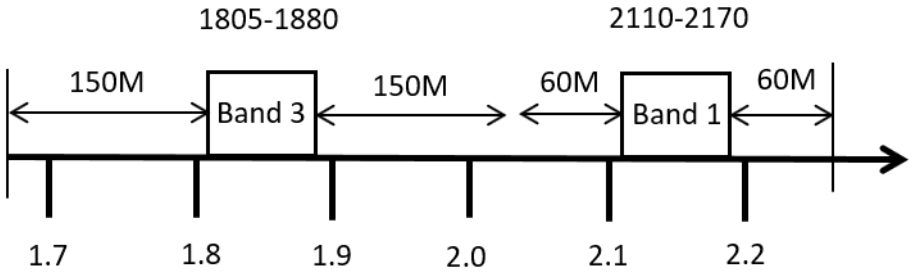


Figure 2.29: Part of the cellular wireless downlink spectrum showing Bands 1 and 3, with bandwidth expansion for DPD. The total bandwidth is about 600 MHz.

converters (ADCs) a few years ago. As a result, the research branch of multi-band DPD proposed new behavioural models to deal with the nonlinearity of each band separately instead of sampling the signals as a full band [3, 4, 49–54].

Existing multi-nonlinear distortion mechanisms lead to significant out-of-band spectral regrowth and in-band error vector magnitude (EVM) degradation. Some significant work [49,51,52] has been proposed to explore multiple-input-multiple-output (or multi-dimension) behavioral models/DPDs for modeling/linearizing cross distortions generated by the interaction of the multi-bands.

Regarding real-time implementation and practical applications of the concurrent multi-band DPD, most of the techniques require multi-parallel feedback loops to instantaneously capture the output signals in each band. It increases the cost of RF components and ADCs in the feedback loop. A tri-band DPD system is shown in Fig. 2.30. In [3] Chao proposed a single feedback solution to reduce the number of feedback loops, shown in Fig. 2.31. Only a single ADC which need to cover the widest bandwidth of multi-band is required in this method. However, considering the dynamic allocation of power and bandwidth in each band, it exhibits risks for practical applications since the output of each band cannot be sampled simultaneously using this method. In [4] Liu proposed another single feedback solution for multi-band using a cascade of mixers, shown in Fig.2.32.

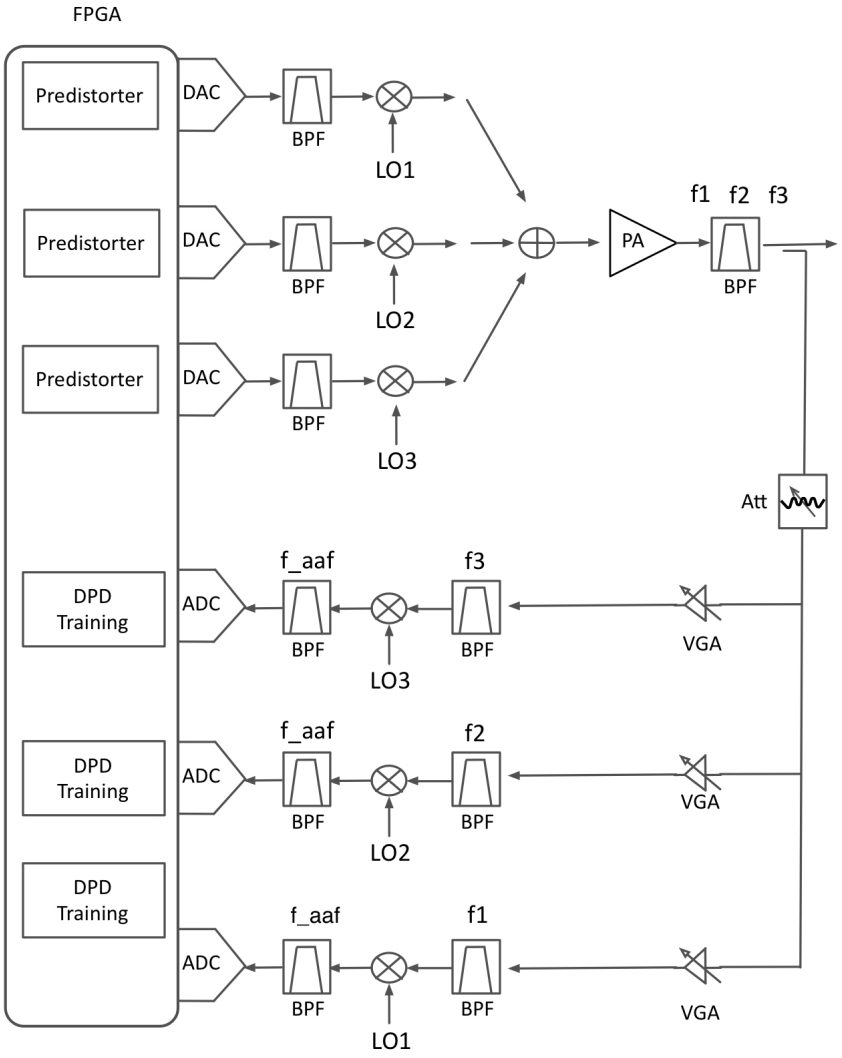


Figure 2.30: Traditional Multi-Band DPD scheme

Similar to [3], only one ADC is required; however, the number of mixers and local oscillators (LOs) are still the same as the traditional structure. Meanwhile, additional complexity is required for time alignment. Furthermore, to achieve the best performance, it requires the feedback signal is sampled as full-scale by the ADC. In this structure, the dynamic range of ADC cannot be optimized for each of bands simultaneously. Due to various power configurations are assigned to different bands, transferring all bands to the same frequency simultaneously by multi-mixers can only guarantee the full-scale sampling for the signal with the highest power.

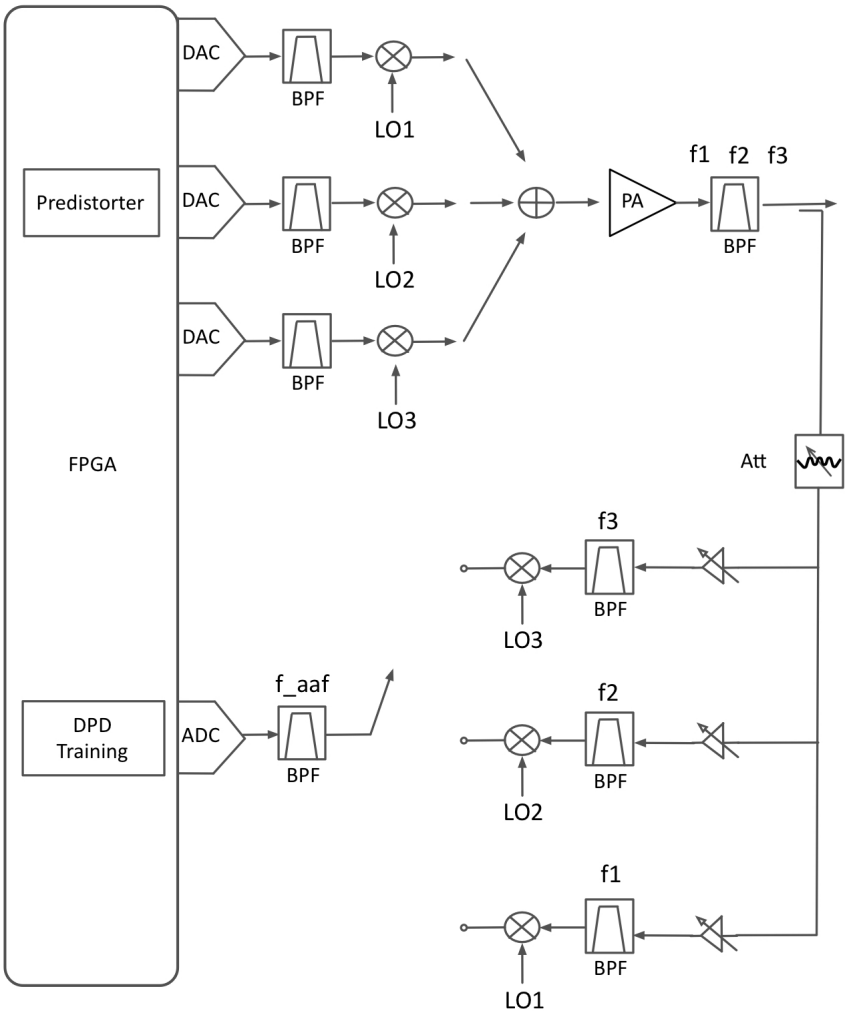


Figure 2.31: Single feedback Multi-band DPD in [3]

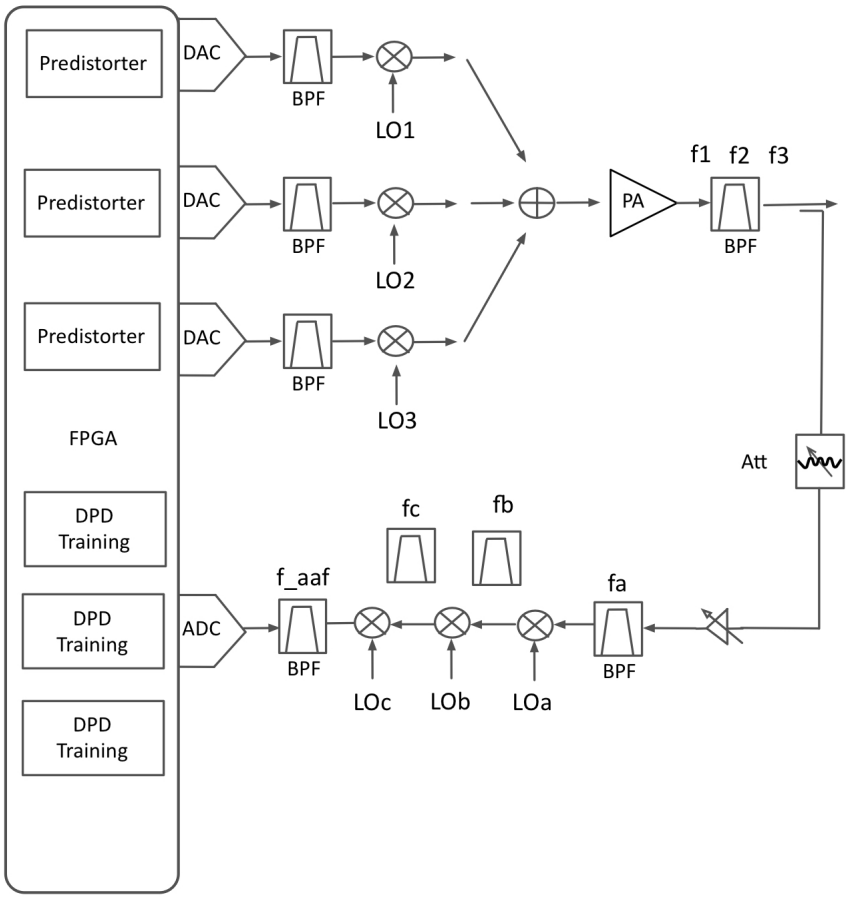
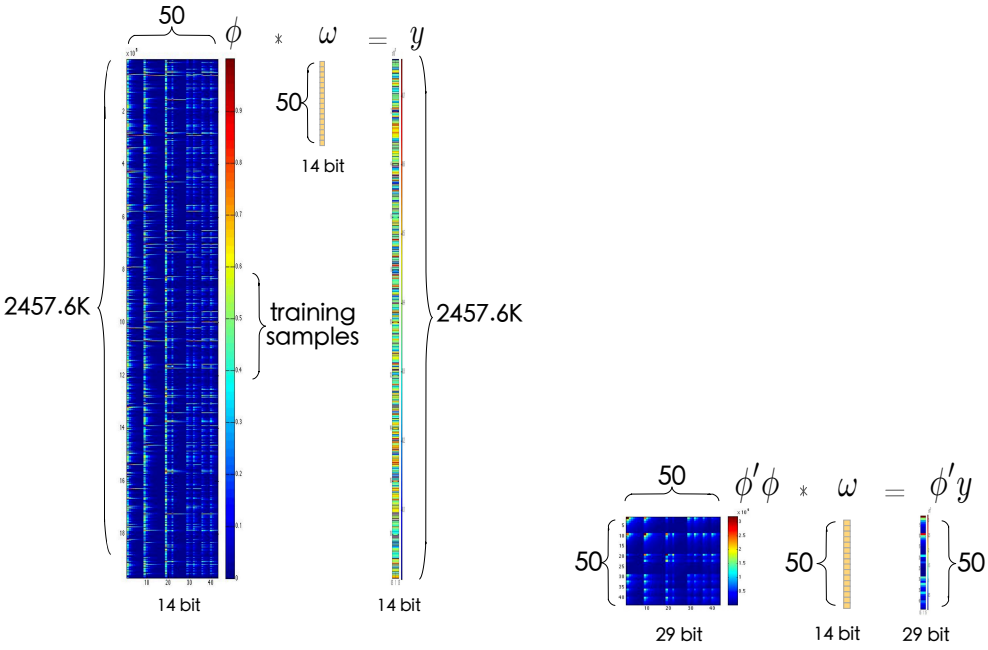


Figure 2.32: Single feedback Multi-band DPD in [4]

Chapter 3

Information Carried by Modulated Signals for Least Square Based Identification

Digital predistortion models based on the Volterra series have been largely employed because of their solid theoretical foundation and straightforward extraction methodology. It is widely accepted that the Volterra-based models are universal approximations, whose structure is linear in the parameters, thus allowing the use of LS solvers to extract the model parameters. To estimate the coefficients of the behavioural model, it requires the training sample set consisting of the system input and system output. As an example, an over-determined equation with 14-bit quantized data, shown in Fig. 3.1a, is firstly constructed using a number of training samples. Then the determined equation with 29-bit quantized data, shown in Fig. 3.1b, is calculated according to the LS formula. It is worth noticing that the a large over-determined equation is compressed into a small determined equation. Moreover, no matter how many training samples are



(a) Over-determined equation (b) Determined equation

Figure 3.1: Over-determined and determined equation using least-square algorithm

in use, the determined equation keeps the same size. As an intuition, with the more training samples, there are more information carried by training samples compressed into the determined equation, which leads to the more accurate of the model estimation.

Thus there is a research question about what is the exact information required by modeling techniques for accurate estimation. The answer is the key to reducing the computational complexity of the coefficients estimation and to reducing the sampling-rate of the observation path. However, some other interesting questions are actually the motivation of this chapter. Do we have to estimate a new set of DPD coefficients if we change the modulation scheme of the transmitted signal? When do we have to find the new set of coefficients, and when can we keep the same coefficients even if another type of modulated signal is transmitted ? This chapter defines the information carried by modulated signals used for least square based algorithms, given a polynomial based behavioural model

with and without memory terms. Then we analysis that how different types of modulated signals effect identification by carrying different information.

3.1 Information carried by Training Samples

Identification method used in DPD systems treats the input and output signal as random and uncertain samples. Little attention is paid to the identification/ adaptation subsystem that most of the time is addressed in the literature by simply solving the least squares (LS) regression applying the Moore-Penrose inverse to extract the DPD parameters. This section revisits the least-square method from the angle of probability information of the training samples in a memory polynomial model.

Without loss of generality, we use a general memory polynomial kernel $\phi_{n,m}(\cdot)$ to illustrate the theory, and for the sake of simplicity we use $\phi(\cdot)$ in this section. This general kernel indicates the relationship between the model basis.

$$\begin{aligned}\phi\big(&\\ &x(n), x^2(n), ..., x^{k_0}(n)\\ &x(n-1), x^2(n-1), ..., x^{k_1}(n-1)\\ &\vdots\\ &x(n-m), x^2(n-m), ..., x^{k_m}(n-m)\\ \big)&\end{aligned} \tag{3.1}$$

where $x(n)$ is the complex input signal, m is the memory depth and k_m is the highest order in the basis with memory length m.

The behavioral model of PAs can be represented as,

$$\phi \cdot \omega = y \tag{3.2}$$

where ϕ is the behavioural model, ω are the coefficients of the behavioural model and y is the output signal of nonlinear PA.

The coefficients can be estimated using the training samples in the direct learning architecture (DLA) [24] or the indirect learning architecture (ILA) [23], shown in Fig. 3.2. The DLA in general has better performance than the ILA, which is more sensitive to the feedback noise because the nonlinear matrix is based on noisy feedback signal [25] [26]. Moreover, in DLA, the behavioural model is constructed only by the input signals in both forward path and the estimation procedure. Subsampling does not work with indirect learning DPD adaption schemes as those methods compare predistorted and postdistorted signals, which are necessarily at the high sample rate of the DPD signal bandwidth [55]. Thus, in this thesis, we select the DLA as the learning structure.

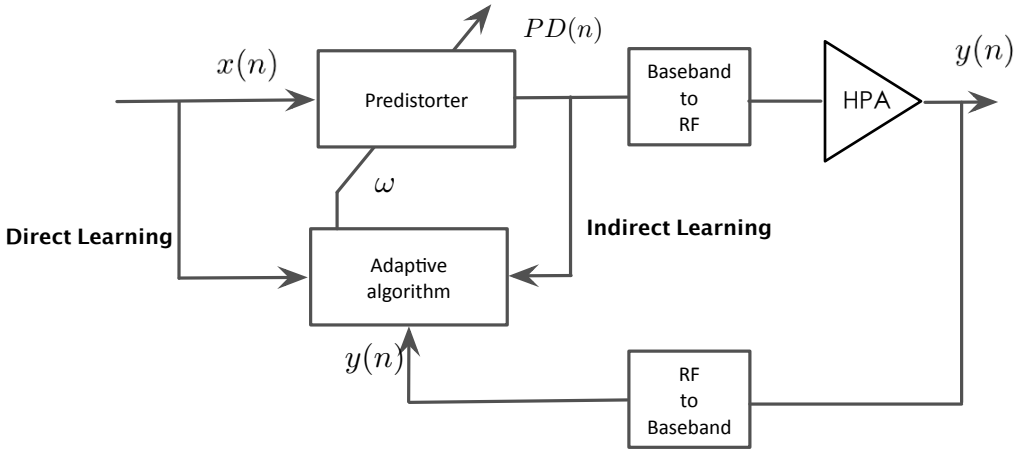


Figure 3.2: Direct learning architecture and indirect learning architecture for coefficients estimation.

It has been well documented that coefficients estimation for such models can

be performed by solving the over-determined equation using Least Square (LS) techniques.

$$\phi^{\mathrm{T}}\phi \cdot \omega = \phi^{\mathrm{T}} y \tag{3.3}$$

In (3.3), $\phi^{\mathrm{T}}\phi$ and $\phi^{\mathrm{T}}y$ are the auto-covariance and cross-covariance matrices respectively. And if we apply expectation calculation function E[] on the above equation, we will have,

$$E[\phi^{\mathrm{T}}\phi \cdot \omega] = E[\phi^{\mathrm{T}} y] \tag{3.4}$$

$$E[\phi^{\mathrm{T}}\phi] \cdot \omega = E[\phi^{\mathrm{T}} y] \tag{3.5}$$

Since modulated signals can be treated as standardized random variables, the covariances equal the correlations [56].

$$R_{\phi\phi} = E[\phi^{\mathrm{T}}\phi] \tag{3.6}$$

$$R_{\phi y} = E[\phi^{\mathrm{T}} y] \tag{3.7}$$

where $R_{\phi\phi}$ is the auto-correlation matrix and $R_{\phi y}$ is the cross-correlation matrix. As a result, the coefficient estimation can be performed using (3.8).

$$\hat{\omega} = R_{\phi\phi}^{-1} \cdot R_{\phi y} \tag{3.8}$$

According to (3.8), we know the coefficients estimated using LS-based algorithms are related to the auto-correlation and cross-correlation matrix. And moreover,

we have

$$R_{\phi\phi} = \int \phi^{\mathrm{T}}\phi \cdot \rho(\phi)\ \mathrm{d}\phi \tag{3.9}$$

$$R_{\phi y} = \iint \phi^{\mathrm{T}} y \cdot \rho(\phi, y)\ \mathrm{d}\phi \mathrm{d}y \tag{3.10}$$

where $\rho(\phi)$ and $\rho(\phi, y_n)$ are the probability distribution function(PDF) and joint probability distribution respectively.

In order to comprise the intermodulation information, the training samples for the DPD are required to be captured with sufficient bandwidth (3-5 times the signal bandwidth). In addition, the number of training samples has to be large enough to carry the accurate statistical information for the nonlinear system [57]. As an intuition, different training samples may lead to different coefficients since the information it carried maybe different. According to (3.9) and (3.10), as shown in Fig. 3.3, the joint probability function $\rho(\phi, y_n)$ can be treated as a projection function for the over-determined equations (3.2) projecting to the determined equation (3.3). The coefficients directly come from the determined equation. It is worth noting no matter how many training samples are used and what's the sampling-rate, the size of the determined equation stays constant. However, different training samples may lead to different joint probability functions of the behavioural model and feedback samples which leads to different determined equation according to (3.9) and (3.10). So we can conclude that the joint probability function $\rho(\phi, y_n)$ is the information carried by the training samples for the coefficients estimation using LS-based algorithms.

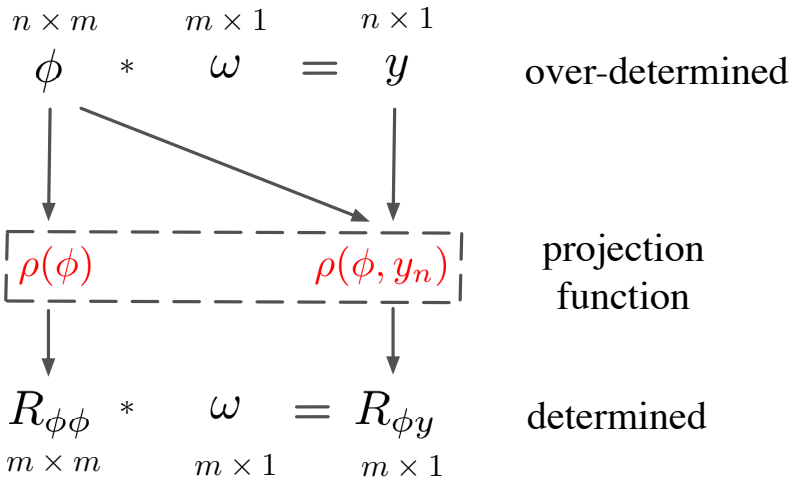


Figure 3.3: The over-determined equation projects to the determined equation according to the joint probability density function

3.2 Joint Probability Information in Nonlinear Systems

3.2.1 Memoryless system

In a memoryless nonlinear system, the general polynomial kernel ϕ becomes

$$\phi\big(\quad x(n),\ x(n)^2,\ ...,\ x(n)^k\quad\big)$$

Extending $\rho(\phi, y_n)$ to the chain rule of probability [58], we obtain:

$$\begin{aligned}
&\rho(x_n,\ {x_n}^2,\ ...,\ {x_n}^k, y_n) \\
&= \rho(x_n) \\
&\quad \cdot \rho(y_n|x_n) \cdot \rho(x_n^2|x_n, y_n) \cdot \rho(x_n^3|x_n, x_n^2, y_n) \\
&\quad \cdots \rho(x_n^k|x_n, x_n^2, \cdots, x_n^{k-1}, y_n)
\end{aligned} \tag{3.11}$$

As for the memoryless system, the output sample y_t at moment t are only

dependent on the input sample x_t at this moment.

$$\rho(y_n|x_n) = 1 \tag{3.12}$$

Furthermore, it is easy to prove that $x_n^2 \cdots x_n^k$ are fixed for a given x_n.

$$\rho(x_n^2|x_n, y_n) = \cdots = \rho(x_n^k|x_n, x_n^2, \cdots, x_n^{k-1}, y_n) = 1 \tag{3.13}$$

Hence for a memoryless system, combining (3.11) with (3.12) and (3.13), we have

$$\begin{aligned} &\rho(x_n,\ {x_n}^2,\ ...,\ {x_n}^k, y_n) \\ &= \rho(x_n) \end{aligned} \tag{3.14}$$

3.2.2 System with memory-effect

Let us consider the chain rule of the joint probability $\rho(\phi, y_n)$ directly (the probability distribution function of memory polynomial and PA output signal).

$$\begin{aligned} &\rho(x_n, {x_n}^2, ..., {x_n}^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n) \\ &= \rho(x_n) \cdot \rho(x_n^2|x_n) \cdots \rho(x_n^k|x_n, x_n^2, \cdots, x_n^{k-1}) \\ &\cdot \rho(x_{n-1}|x_n, \cdots) \cdots \rho(x_{n-1}^{k1}|x_{n-1}, \cdots) \cdots \\ &\cdot \rho(x_{n-m}|x_n, x_{n-1}, ..., x_{n-m+1}, \cdots) \cdots \rho(x_{n-m}^{k_m}|x_{n-m}, ...) \\ &\cdot \rho(y_n|x_n, x_{n-1}, x_{n-2}, ..., x_{n-m}, ...) \end{aligned} \tag{3.15}$$

One should notice that (3.12) does not hold in the case of a system with memory since the output at time t is not only determined by the input signal at time t but at time $t-1$ to $t-m$. And (3.13) still holds for a system with memory, so

(3.15) can be rewritten as follows:

$$\begin{aligned}
&\rho(x_n, {x_n}^2, ..., {x_n}^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n) \\
&= \rho(x_n) && \text{(3.16a)} \\
&\cdot \rho(x_{n-1}|x_n) \cdot \rho(x_{n-2}|x_n, x_{n-1}) \\
&\quad ...\rho(x_{n-m}|x_n, x_{n-1}, ..., x_{n-m+1}) && \text{(3.16b)} \\
&\cdot \rho(y_n|x_n, x_{n-1}, x_{n-2}, ..., x_{n-m}) && \text{(3.16c)}
\end{aligned}$$

If the behavioural model of DPD is comprised of sufficient memory taps, then the output sample of the nonlinear system will be fixed given the sequence $x_n, x_{n-1}, ..., x_{n-m}$. Thus, given a behavioural model with sufficient number of memory taps, we have

$$\rho(y_n|x_n, x_{n-1}, x_{n-2}, ..., x_{n-m}) = 1 \tag{3.17}$$

Then the joint probability can be written as

$$\begin{aligned}
&\rho(x_n, {x_n}^2, ..., {x_n}^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n) \\
&= \rho(x_n) && \text{(3.18a)} \\
&\cdot \rho(x_{n-1}|x_n) \cdot \rho(x_{n-2}|x_n, x_{n-1}) \\
&\quad ...\rho(x_{n-m}|x_n, x_{n-1}, ..., x_{n-m+1}) && \text{(3.18b)}
\end{aligned}$$

Now consider the non-bandlimited signal. The time-discrete non-bandlimited samples are independent random variables with I/Q being gaussian distributed [59]. As a result, in terms of the non-bandlimited signal, the conditional proba-

bility (3.22b) becomes the marginal probability (3.22a).

$$\rho(x_{n-p}|x_n) = \rho(x_{n-p}) = \rho(x_n), \qquad p \geq 1 \tag{3.19}$$

Assume that the relative bandwidth ratio of the bandlimited signal used in DPD system is B which is about 3 to 5 in practical. So we can simplify the conditional probability as

$$\rho(x_{n-p}|x_n) = \rho(x_n),, \qquad p \geq B \tag{3.20}$$

, where B is the ratio between the linearisation bandwidth and the bandwidth of the transmitted signal.

Since modulated signals satisfy stationary property, the probability function is not a time-varying function.

$$\rho(x_{n-m}|x_n, x_{n-1}, ..., x_{n-m+1}) = \rho(x_{n-m-1}|x_{n-1}, x_{n-2}, ..., x_{n-m}) \tag{3.21}$$

Substitute (3.21) and (3.20) into (3.18), we can simplify the joint probability.

$$\begin{aligned}
&\rho(x_n, {x_n}^2, ..., {x_n}^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n) \\
&= \rho(x_n) && (3.22a) \\
&\cdot \rho(x_{n-1}|x_n) \cdot \rho(x_{n-2}|x_n, x_{n-1}), ..., \rho(x_{n-p}|x_n, x_{n-1}, ..., x_{n-p+1}) && (3.22b) \\
&p = min\{B-1, m\}
\end{aligned}$$

where B is the ratio between the linearisation bandwidth and the bandwidth of the transmitted signal, m is the length of memory taps in the behavioural model.

In brief, in terms of a nonlinear system with memory-effect, as proven above,

given a behavioural model with sufficient memory taps, the joint probability $\rho(\phi, y_n)$ can be rewritten in (3.22). To further simply the joint probability in realizable DPD system, for example, if we assume that the linearisation bandwidth is 3 times wider than the signal bandwidth ($B = 3$) and the length of memory taps is more than 3, the joint probability function can be written as

$$\rho(x_n, {x_n}^2, ..., {x_n}^{k_0}, x_{n-1}, x_{n-1}^2, ..., x_{n-1}^{k_1}, ..., x_{n-m}^{k_m}, y_n)$$

$$= \rho(x_n) \tag{3.23a}$$

$$\cdot \rho(x_{n-1}|x_n) \cdot \rho(x_{n-2}|x_n, x_{n-1}) \tag{3.23b}$$

Using LS based algorithms, the coefficients are obtained from a determined equation (3.3). As for standardized random variables, the determined equation can be treated as the auto and cross-correlation matrix in (3.8). The correlation matrices are related to the probability density function as shown in (3.9) and (3.10). As a result, the joint probability $\rho(\phi, y_n)$ is the information carried by samples. In this section, we have shrunk the definition of information carried by training samples to (3.23). The probability information is now only determined by the statistic of the transmitted signal. In the first time, a mean is proposed to further analysis statistic features of the LS algorithm while memory-effects are considered in the model. In the next section, it will connect specific statistical properties of telecommunication signals with this definition which leads to further conclusions.

3.3 Analysis of the Information Carried by Training Samples

The memory-polynomial based behavioural model cannot perfectly fit a PA. It is widely used for digital predistortion system as it reduces the calculation complexity by avoiding physical level effects. Thus, unlike the physics-based modeling, coefficients of a behavioural model are only optimized for the current condition of the PA. The condition of a PA includes the temperature, gain, output power level and transmitted signal. As a rule of thumb, the accuracy of the behavioural modeling is expected to be degraded if the condition of the PA changes but coefficients are not updated. Coefficients of the model are estimated only based on training samples which are a part of samples from the transmitted signal. Intuitively, the low-rate or a small number of training samples may lead to the degradation of modeling/DPD accuracy. On the other hand, transmitting a different signal may also require updating or re-training the coefficient of DPD. In this section by analyzing the information carried by training samples which are simplified in (3.23), it proposes that PAPR value and bandwidth of original signal determine the information used for LS estimation. Thus, various training sample sets with different sampling rates, different numbers of samples or different numbers of carriers could lead to the same estimated coefficients once given the same PAPR and signal bandwidth. Furthermore, an simulation example shows that two different training sample sets, which are 3-carrier LTE and 5-carrier LTE signal, can share the same estimated coefficients if they have the same PAPR and signal bandwidth. The analysis in this section also indicates the feasibility of estimating the coefficient of DPD from the low sampling-rate and the small number of training samples by reserving the information in (3.23).

In modern wireless communication systems, we use complex baseband signals

in-phase(I) and quadrature(Q) components. Since the probability distributions of I and Q components of the modulated signal are independent, the PDF of the baseband signal can be written as below:

$$\rho(x_n) = \rho(x_n^I) \cdot \rho(x_n^Q) \cdot \rho(|x_n|) \tag{3.24}$$

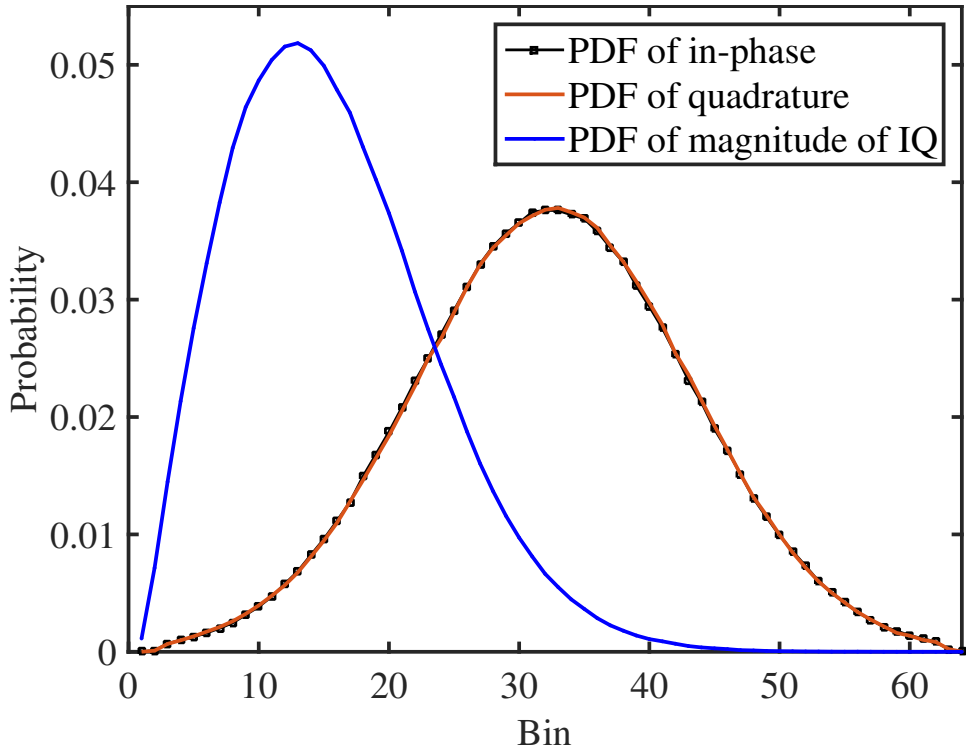


Figure 3.4: Probability Distribution of a 20 MHz LTE Signal

The I and Q components are typically Gaussian-distributed resulting in that the absolute value $|x_n|$ of the complex signal is Rayleigh-distributed [60], as shown in Fig. 3.4. Moreover, given modulated I/Q signals, the Rayleigh-distribution function has only one variable, i.e., the peak-to-average power ratio(PAPR). Fig. 3.5 shows different PDFs of normalized LTE signal with different PAPRs. A 1-carrier LTE signal and a 4-carrier LTE signal have similar PDFs since they have similar PAPR value. The two 1-carrier LTE signals with different PAPR have different PDFs. Although modern modulated signals are random in

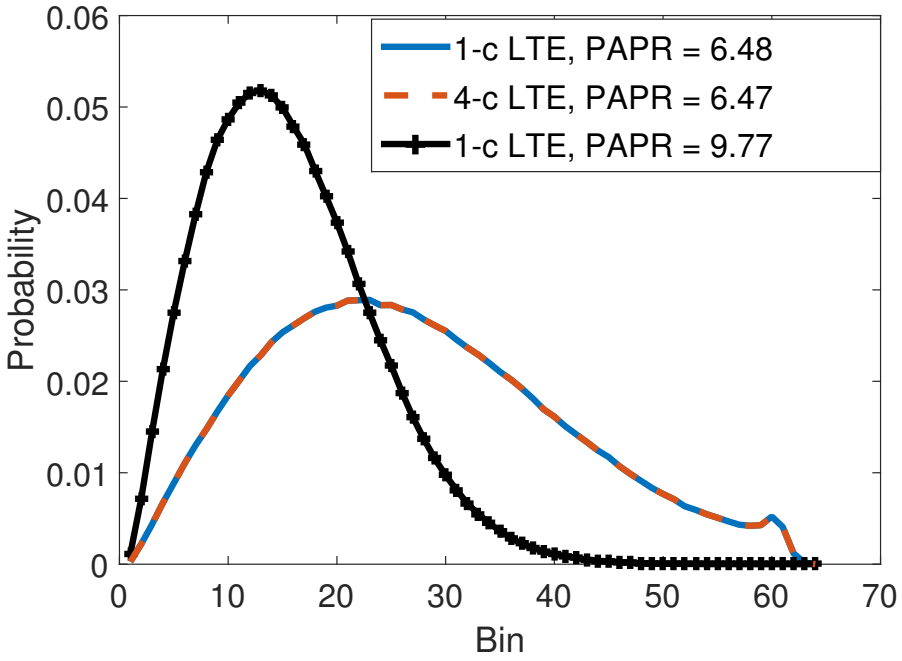


Figure 3.5: Probability distribution of signals with different PAR

time domain, the term (3.22a) is only determined by the PAPR value given the normalized power.

The term (3.22b) depends on the relative bandwidth ratio which is the ratio of the linearisation bandwidth(instantaneous bandwidth) to the signal bandwidth (occupied bandwidth). The wider the bandwidth, the wider the range of terms (3.22b). As shown in Fig. 3.6, given the same data rate of 208 MSPS, the red lines indicates the range of $\rho(x_{n-1}|x_n)$ in 80 MHz 4-carriers LTE, 20 MHz 1-carrier LTE, 20 MHz 4-carriers GSM and single tone respectively. The wider signal bandwidth, the bigger range of $\rho(x_{n-1})$ given a $|x_n|$ value (shown as red dashed line of 80MHz signal and 20MHz signal). On the other hand, with the same signal bandwidth, 1-carrier LTE signal and 4-carrier GSM signal occupy the same distribution in Fig. 3.6. From this figure, it indicates that the signal bandwidth determines (3.22b).

There is a question at the begin of this chapter: do we have to find new coef-

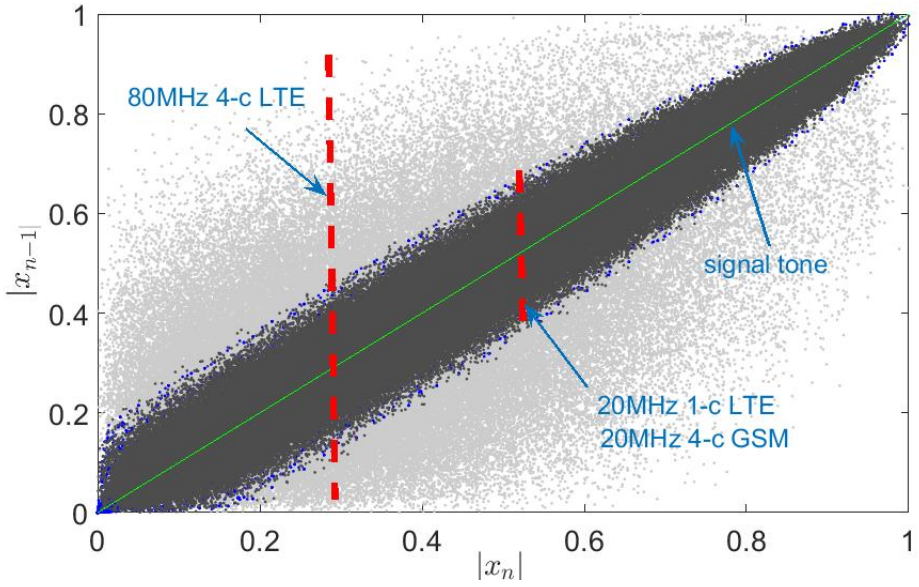


Figure 3.6: the Range of x_{n-1} given x_n. The grey point is 80 MHz 4-carriers LTE signal, the black point is 20 MHz 1-carrier LTE signal, the blue point is 20 MHz 4-carriers GSM signal and the green point is single tone.

ficients of DPD if the transmitted signal changes? It depends on whether the bandwidth and PAPR of transmitted signal change. The coefficients estimation is carried out using the information carried by transmitted signals. If the bandwidth and PAPR value of the transmitted signal keep the same, the DPD coefficients suit the new transmitted signal. In contrast, the modulated signal with different signal bandwidth and PAPR carries different information. Thus for this condition, DPD coefficients are required to be updated. There is a toy example to demonstrate the effect of the information on the coefficients estimation. An ARCTAN PA model, introduced in [61], is selected to generate the nonlinearity and memory-effect. A 3-carrier 100MHz-bandwidth 7dB-PAPR LTE signal is generated as the transmitted signal. The memory polynomial model with 100 coefficients (10 orders and 10 delay taps) is selected as the behavioural model of the DPD. The spectrum result and the AMAM result with and without DPD are shown in Fig. 3.7 and Fig. 3.8. Then, a 5-carrier 100MHz-bandwidth 7dB-PAPR LTE signal is used as the transmitted signal. The 5-carrier LTE signal

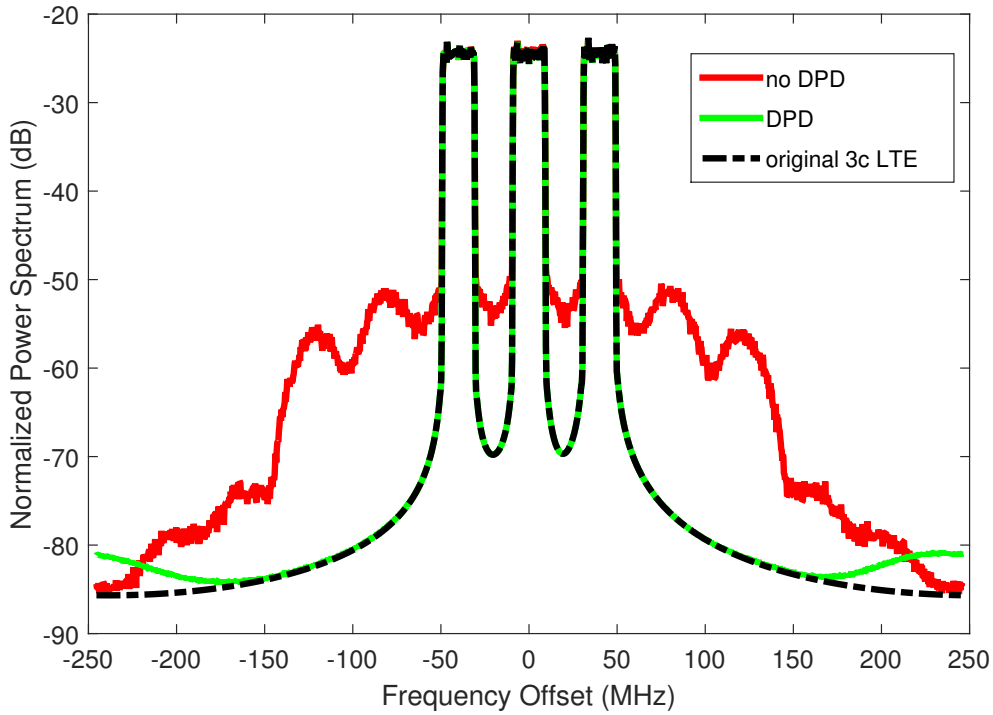


Figure 3.7: 3 carriers 100MHz 7.0dB PAPR LTE signal with and without DPD

has the same signal bandwidth and PAPR value as the 3-carrier LTE signal. Thus, the DPD coefficients calculated based on the 3-carrier signal can be directly shared to the new 5-carrier LTE signal. The spectrum result using the shared coefficients from the 3-carrier DPD is shown in Fig. 3.9. In contrast, a 2-carrier 120MHz-bandwidth 9dB-PAPR LTE signal cannot use the DPD coefficients estimated from the 3-carrier LTE signal because the signal bandwidth and PAPR is changed. Fig. 3.10 shows that DPD collapses by using the coefficients which is estimated from the signal carrying different probability information.

To summarize the take-away message, in a memoryless nonlinear system, the coefficients of the polynomial type behavioural models are only related to the term (3.14) (or the term (3.22a)) which is called the nonlinearity term and is determined by the PAPR value of signals. In a nonlinear system (power amplifiers) with memory effects, the coefficients of polynomial type behavioural

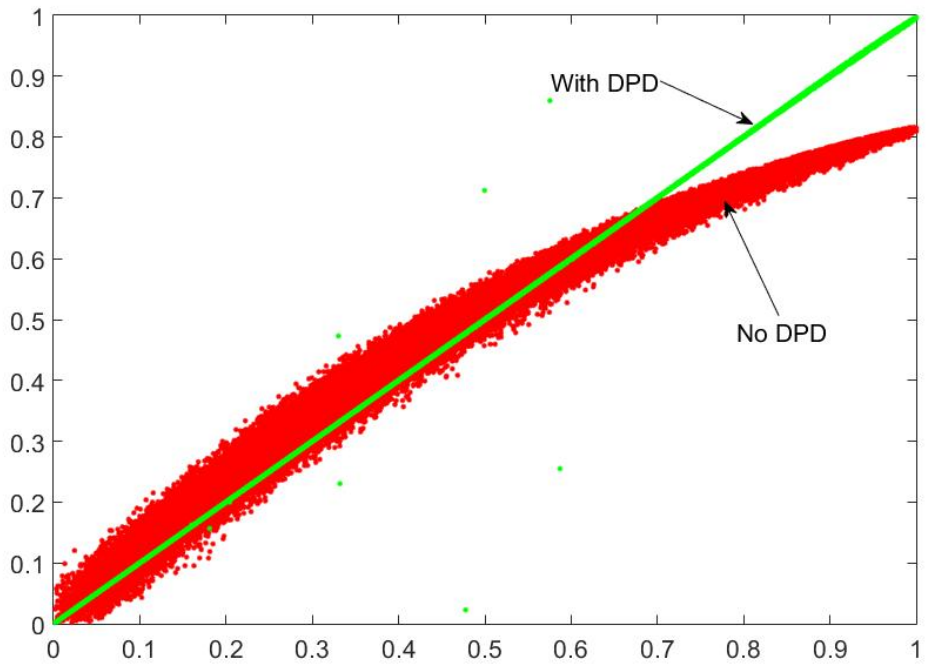


Figure 3.8: AMAM of the PA model with and without DPD

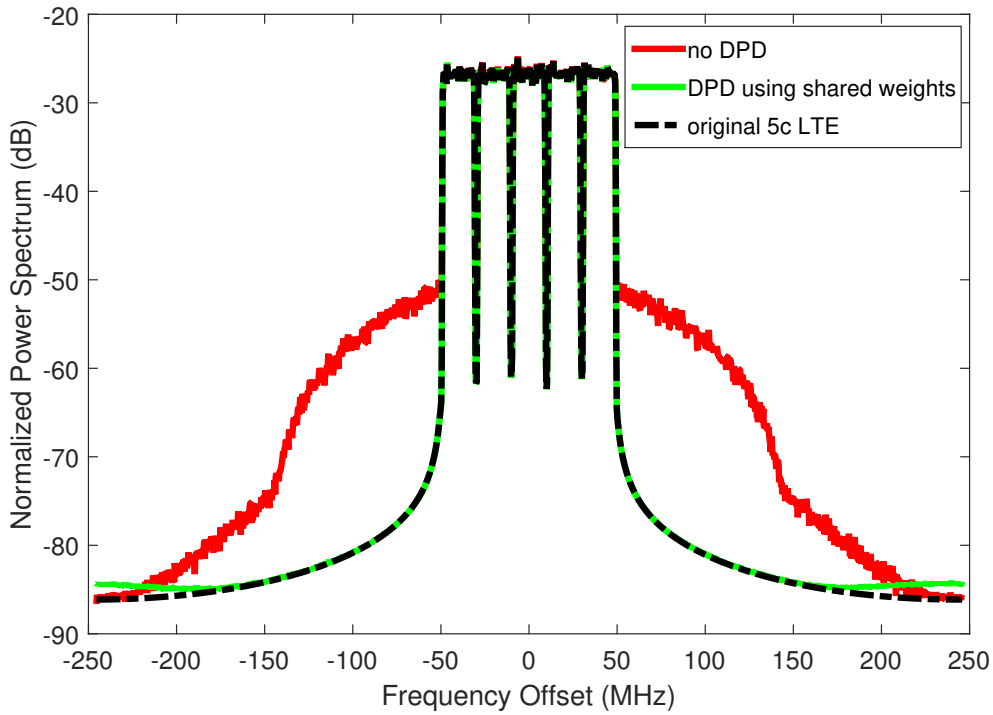


Figure 3.9: 5 carriers 100MHz 7.0dB PAPR LTE DPD using the weights calculated from a 3-C LTE signal

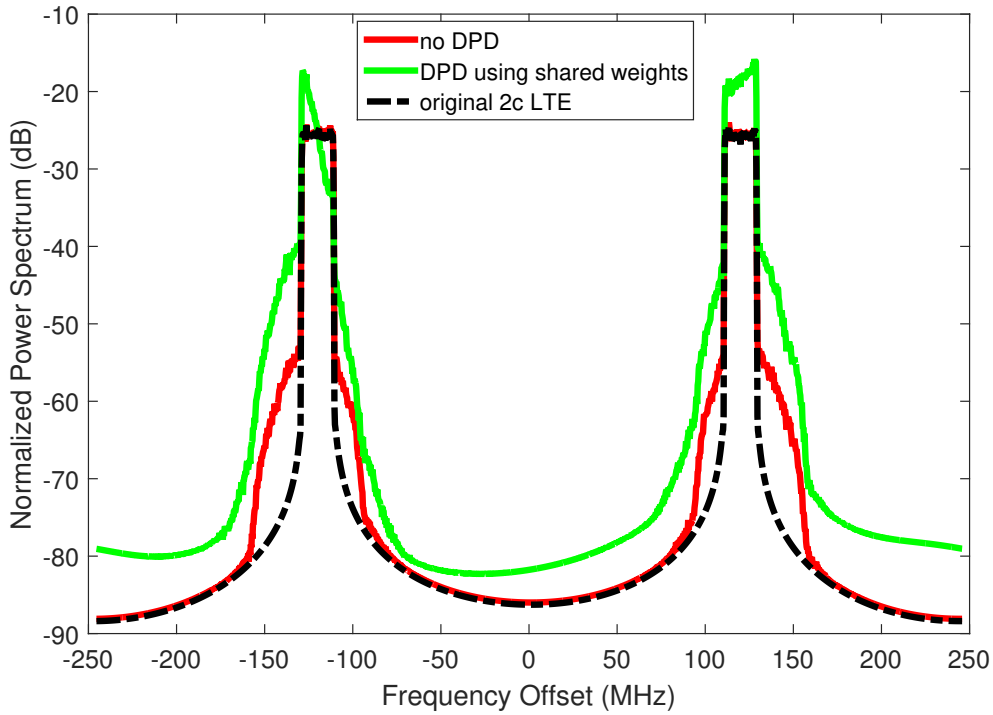


Figure 3.10: 2 carriers 120MHz 9.0dB PAPR LTE DPD using the weights calculated from a 3-C LTE signal

models are related to both the nonlinearity term (3.22a) and the bandwidth term (3.22b). Furthermore, the nonlinearity term is only determined by the PAPR value of modulated signals and the bandwidth term is only determined by the relative bandwidth ratio of the modulated signal. It is worth noting that the information of the training samples for modeling in (3.22) is only comprised of input signal x_n. Even though the modulated signal is random in the time domain, the joint probability information comprised of the modulated signal is stable and predictable, given PAPR and signal bandwidth. More importantly, various training sample sets with different sampling rates, different numbers of samples or different numbers of carriers could lead to the same estimated coefficients once given the same PAPR and signal bandwidth. This provides the theory feasibility of reducing the number of training samples and reducing the feedback sampling-rate which are discussed in Chapter 4 and Chapter 5.

Chapter 4

Compressed Training Samples for Behavioural Modeling and Digital Predistortion

4.1 Introduction

The characteristic behavior of nonlinear power amplifiers can be efficiently and accurately represented using behavioral models. The two main procedures which are carried out using behavioral models are the extraction of model parameters and the calculation of model outputs. It is generally accepted that a model with good generalization capabilities can be extracted from common mobile telecommunication standards using the discrete complex envelope representations of the input and output signals. Calculating the model parameters for discrete signals in some digital signal processing hardware often requires that a very large number of consecutive samples are recorded to improve the generalisation capabilities. Further to this, with the proliferation of high efficiency PA architectures

such as Doherty and outphasing PAs for multi-band, multi-standard, even more samples are needed to extract a model which will encompass all characteristics of the PA.

In the literature, most of the DPD solutions are proposed for medium or high output power PAs, which are normally employed in middle-to-large-size macro-cell base stations. For small base stations, mobile handsets and the massive MIMO transmitters, low-power RF amplifiers still suffer from lower efficiency because of the use of power back-off to control the inherent nonlinear distortions. On the other hand, since the signal bandwidth gets wider, meanwhile flexible or non-contiguous spectrum access is introduced, the complexity of the DPD increases and the power consumption added in the digital domain by the DPD itself becomes higher and higher. As a result, the DPD added power consumption becomes more sensitive to the overall efficiency, to the extent that it might be close to, or even surpass the energy saving achieved from using a more efficient PA. Therefore, a big challenge is how to design the low-power and real-time DPD without sacrificing the linearization performance.

One strategy commonly employed to reduce the computational complexity of large matrix computations required to train a behavioral model, is to reduce the number of model parameters [62] [63] [64] [65] [66]. This is an effective means of reducing both the extraction time and the calculation of model outputs. However it does not guarantee the generalisation properties of the model. It is still required to have a good training signal which encompasses the characteristics of the PA to train the reduced dimension model.

In the case of the RF PA behavioural modeling and DPD, the typical selection process for behavioral model training data involves the selection of consecutive data samples which include key samples such as the maximum magnitude and

maximum rate of change. Since modern communication standard signals are not truly periodic, it is not possible to simply select a short section of consecutive data to represent the whole dataset. Guan *et al* proposed an algorithm to use multiple short training data to build multiple small matrices for behavioural modeling [57]. However, this method does not reduce the total number of training samples and cannot guarantee the robustness for a nonlinear system with memory.

This chapter presents an algorithm which uses the probability information of the input signal to inform the selection of a compressed training dataset for RF PA behavioural model extraction. The proposed algorithm can dramatically reduce the number of training samples. The accuracy of this algorithm is validated by extraction of behavioural models using a large dataset of consecutive samples and a reduced training dataset determined using the proposed algorithm. A significant reduction in computational complexity and faster execution time is achieved with the new approach.

4.2 Computation Complexity of DPD Training

The number of calculations per second the DPD must accomplish can be in the teraflop range [67]. To estimate the coefficients of the behavioural model/DPD, LS-based algorithms are normally employed. The LS algorithm is a statistical regression-based method that finds a coefficients of the model by minimizing the sum of the squares error between the expected and observed data. Since it is not practical to involve all the data in the coefficients estimation, training samples, including the input and output samples, are required to represent the whole data set. The LS algorithm is known as the best linear unbiased estimator of the coefficients, and the global solution can be estimated when the errors

among each set of training samples are uncorrelated. In the DPD system, the output signal is distorted but is still correlated to the input signal, thus the observation errors are not uncorrelated. To reduce the influence of the errors in the training procedure, the LS estimator requires a large number of training sequences to derive the best approximation of the DPD coefficients which is normally a couple of thousands of samples for practical fixed-point DPD application. Since matrix-matrix multiplication and inversion are very complex to operate, LS with large-size matrices are very expensive to implement. Furthermore, large matrix operations are also time consuming [16], which may cause problems in some applications such as a system that requires real-time or fast coefficients adaptation [57].

The coefficients of the behavioural model/DPD can be estimated by

$$\omega = (\phi^{\mathrm{T}}\phi)^{-1} \cdot \phi^{\mathrm{T}} y \tag{4.1}$$

The LS estimation seems to be simple and straightforward. It requires at least $N \cdot (K(K+1)/2+1)$ to construct the behavioral model ϕ, where N is the number of feedback samples and the K indicates the number of DPD coefficients. It requires $K^3/2 + K^2 - 5K/2$ for Gauss-Jordan method to solve the equation (4.1), where K indicates the number of DPD coefficients.

According to [57], the total number of complex multiplication operations to be conducted in (4.1) is

$$O_{N,K} = N \times K^2 \times 2 + N \times K + K^3 - 5K/2 + N \tag{4.2}$$

where N is the number of training samples, K is the number of coefficients. For example in a typical case, given 16K samples and 100 coefficients, it requires at

least 322 million multiplication operations to implement the LS algorithm.

4.3 Traditional Training Samples Selection

In the DPD training process, the training samples should be selected to represent the nonlinear system. There are hundreds of millions of samples per second in the baseband processing of DPD systems, for example the data rate of the LTE DPD can be any multiple of 122.88 megasamples per second (MSPS). Thousands of consecutive samples, typically from 8K to 16K, are normally used in a DPD system [68] [69]. The feasibility of using consecutive training samples to represent the whole samples set (hundreds million of samples per second) comes from the stationary property of modulated signals.

The modern telecommunications waveforms can be approximated to a wide-sense cyclostationary process [59]. The distribution for any given set of samples will be time-invariant over a time shift nT as described. Thus, the probability distribution of a consecutive samples can be approximated to the probability distribution of the whole data set. This is why selecting a consecutive data as training samples can accurately represent the information of the transmitted signal in DPD estimation.

$$[X_1, \ldots, X_T] \sim [X_{1+nT}, \ldots X_{T+nT}] \qquad \forall n \tag{4.3}$$

Specifically, in the case of a nonlinear system with weak memory-effect, providing there are sufficient consecutive samples in a reduced dataset to satisfy (4.3), the PDF of this reduced consecutive training dataset will be similar to that of the full dataset. Thus, selecting a consecutive modulated training dataset is sufficient to extract the accurate model in a memoryless system. Fig. 4.1 shows

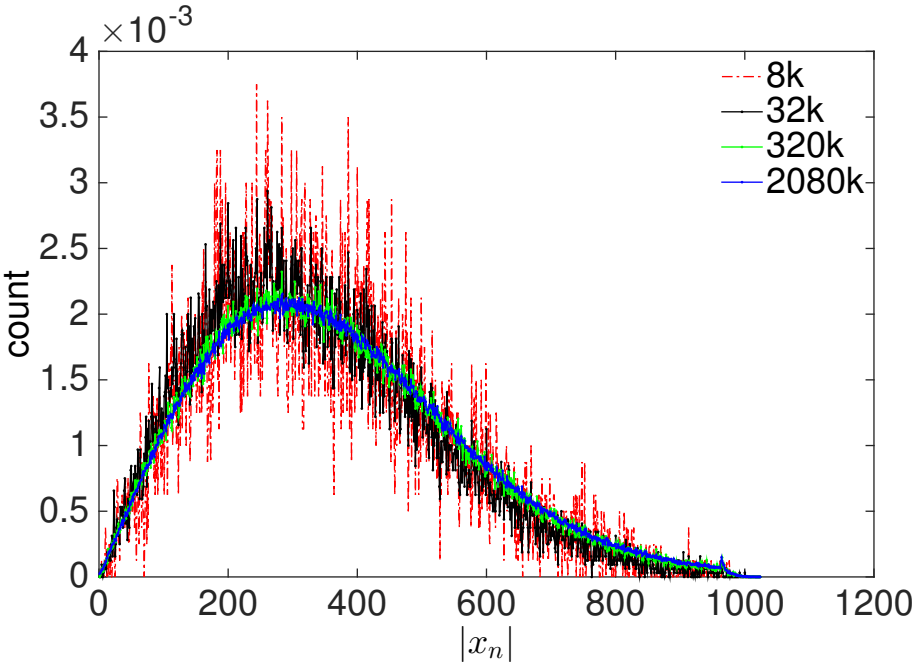


Figure 4.1: Normalized Probability Distribution of LTE Signal with Different Number of Consecutive Samples

the distribution of a 4-carrier LTE signal with different number of consecutive samples. The fundamental shape of all PDF are consistent even with different numbers of samples.

In terms of nonlinearity, the efficiency of using a selected set of consecutive training samples in modelling highly depends on how close the statistical properties of the given signal are to a wide-sense stationary process. It leads to an increase of required training samples for multi-standard signals. On the other hand, in terms of a system with memory, the training samples have to satisfy not only (3.14), but also (3.16) which includes the characteristics due to the signal bandwidth and characteristics due to memory effects. So the more complicated memory-effect mechanisms present in new high efficiency PA architectures, such as Doherty, LINC, ET, EER, will require more consecutive samples to guarantee (3.16c). Similarly for wider bandwidth signals, (3.16b) becomes harder to train

for. As shown in Fig. 3.6, the red lines indicates the range of $\rho(x_{n-1}|x_n)$ in 100MHz and 20MHz LTE respectively.

Since the spectrum of the output of linearized PA is strictly required to meet the 3GPP requirement, DPD system has to precisely model the nonlinear system with memory-effects. For example, -50 dBc ACLR and at least -40 dB NMSE is preferred in some practical DPD systems. -40 dB NMSE indicates 0.01% error which is much lower than many of identification or modeling applications. Moreover the coefficients estimation of practical DPD systems is usually performed in fixed-point as the FPGA or DSP devices lack floating-point multipliers. As a result, it requires thousands of consecutive training samples to accurately estimate the DPD coefficients. However, LS estimation treats the training samples as random and uncertain data. So every samples is equally accumulated to construct the determined equation. It is inefficient to consecutively select a large number of training samples since the distribution of samples is predictable given the bandwidth and PAPR of signal.

4.4 Compressed Training Sample

Firstly we assume there are an over-determined equations with 2 million 14-bit LTE samples and K 14-bit coefficients of memory polynomials. Then by using LS algorithm, the information of the over-determined system is compressed into the correlation space. It is worth noticing that no matter how many training samples are used in the optimization algorithm, (3.3) is always a K by K determined equation, which contain 29-bit compressed information. In Chapter 3, it proposes that the responsibility of training samples is to carry the nonlinearity response (PA output samples) and the training information (3.23). The information used for LS estimation is determined by two terms: the nonlinear-

ity term (3.23a) and the bandwidth term (3.23b). As mentioned in section 4.3, with respect to tradition training samples selection, the joint probability information of a part of consecutive training samples can be approximate to that of the whole dataset because of the cyclostationary property of modulated signals (time-invariant probability). However, it is low efficient to select consecutive samples as it requires a large number of samples to accumulate the accurate PDF. On the other hand, given the PAPR and signal bandwidth, the nonlinearity term (3.23a) and the bandwidth term (3.23b) can be determined before transmitting the signal. Simply selecting the minimum number of samples to carry the known probability distribution could significantly reduce the required number of training samples which leads to a reduction of complexity of the DPD training. Thus, we design a training samples selection method to obtain the same auto-correlation and cross-correlation matrix in (3.8) by involving the minimal samples, rather than roughly using a large section of consecutive samples.

Procedure:

1: Choose a resolution for the histogram;

2: Generate histogram from the whole input signal dataset based on $|x_n|$;

3: Determine the factor to reduce the length of the dataset by;

4: Obtain a target histogram by proportionally reducing each histogram bin size according to the factor, as shown in Fig. 4.6;

5: i) Select a random sample x_j and identify the $|x_j|$ bin based on its magnitude; ii) If the count in $|x_j|$ bin is less than the count in $|x_j|$ bin of the target histogram, keep the sample. Otherwise, disregard the sample;

6: Repeat step 5 until all target histogram bins are full;

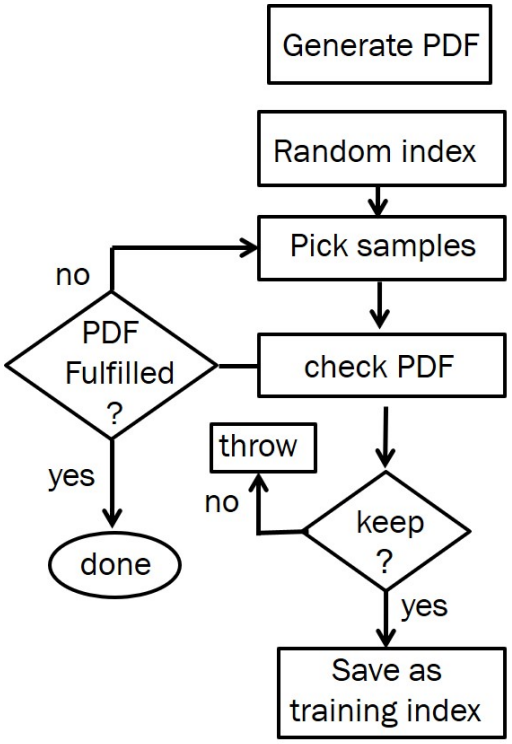


Figure 4.2: Brief flowchart of the proposed sample selection method

7: For each selected sample x_j, we also keep the previous samples $(x_j, x_{j-1}, \cdots, x_{j-m}, y_j)$;

8: Use the selected training samples to calculate the coefficients.

Fig. 4.3 indicates the training samples selected by the traditional procedure and the proposed procedure. The random selection of samples ensures that the terms (3.16b) and (3.16c) can be covered correctly since the various distributions of the whole dataset can be conserved by random selection. Fig. 4.5 shows that the selected samples follow the same $\rho(x_{n-1}|x_n)$ distribution of the whole dataset. The rescaled (normalized) histogram is appropriate to verify the accuracy of probability [70]. Meanwhile because the target histogram is the scaling down histogram of the whole dataset, the information is fully reserved, as shown in Fig. 4.6.

Because the information of (3.16) is reserved, $R_{\phi\phi}$ and $R_{\phi y}$ in (3.8) are guar-

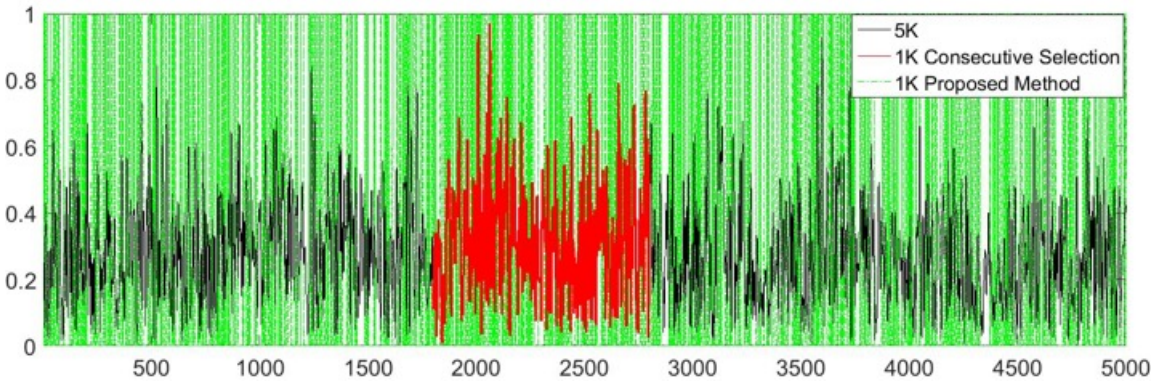


Figure 4.3: Traditional selection and proposed selection

anteed to yield a accurate determined equations. As mentioned in section 4.3, the feasibility of using consecutive thousands of training samples to represent the whole samples set (hundreds million of samples per second) comes from the stationary property of modulated signals. By means of the proposed samples selection method, the type of signal is not restricted to stationary signals but any signals with known PDF. The number of training samples, which carry the necessary information to obtain the accurate determined equation, can be dramatically reduced.

4.5 Experimental Tests of Behavioural Modeling

To validate the performance of the proposed compressed training signal dataset, an experimental testbench was assembled. The hardware testbench comprised of a signal generator(Rohde & Schwarz SMU200A), a 40W Doherty power amplifier and a spectrum analyser(Rohde & Schwarz FSQ). The frequency spectrum of the 4-carrier WCDMA input and PA output are shown in Fig. 4.7. The 14-bit 16K datasets recorded by the FSQ were used to extract the GMP model [21] in (5.9) with 40 coefficients. Then the coefficients of the model are calculated by means

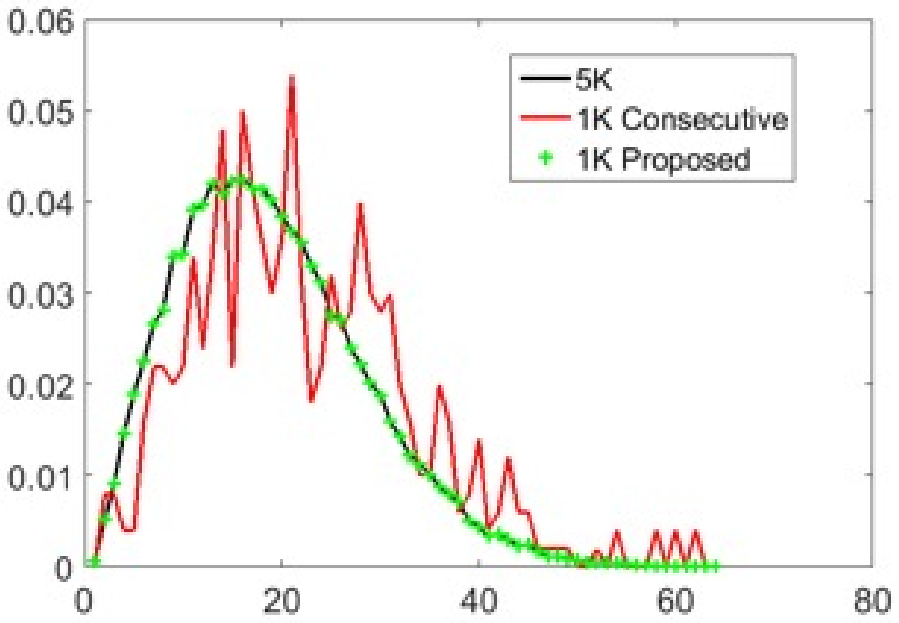


Figure 4.4: The distribution of the training samples selected by traditional and proposed procedure.

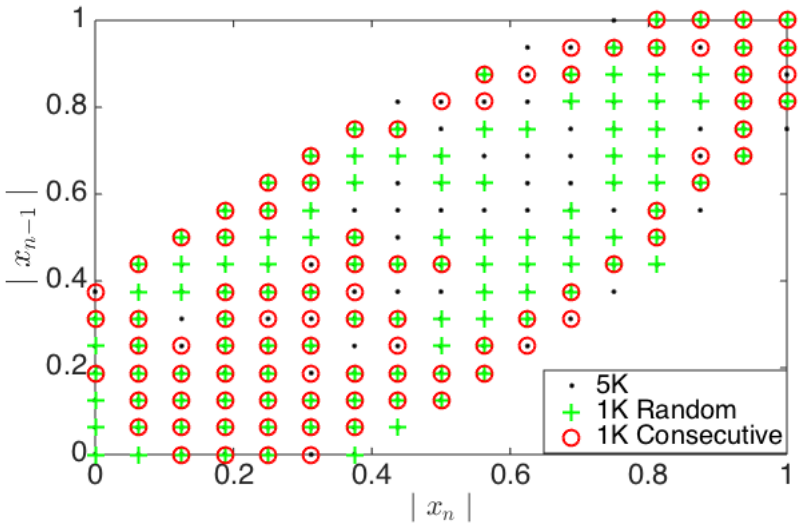


Figure 4.5: The distribution of x_{n-1} for a given x_n in 5-bit resolution. The grey point is the whole dataset of 5k multi-tone samples. The red ring is 1K consecutive samples including maximum magnitude. And the green cross is 1K selected samples from random dataset.

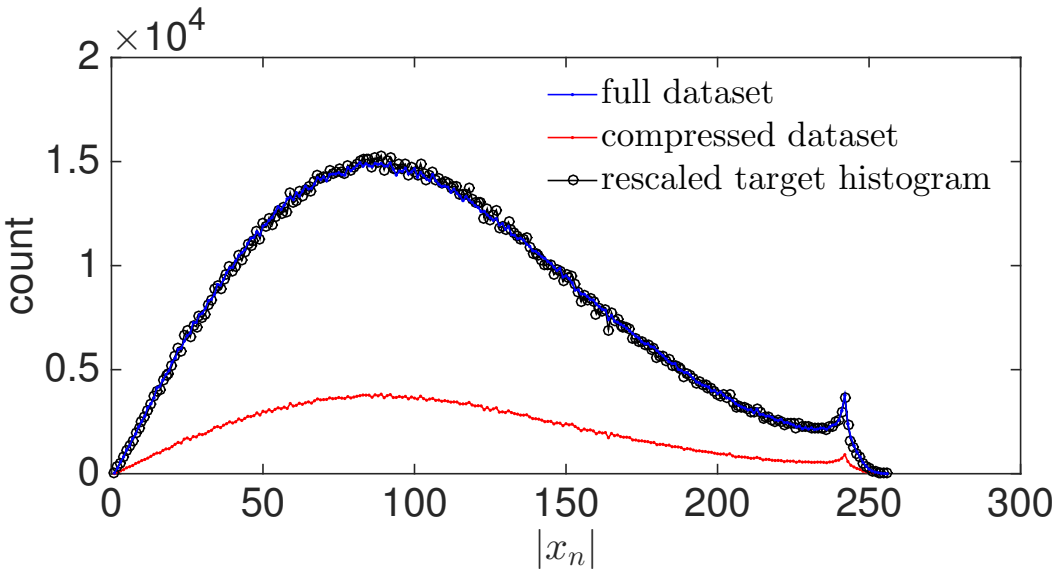


Figure 4.6: Histogram for full dataset and compressed training dataset

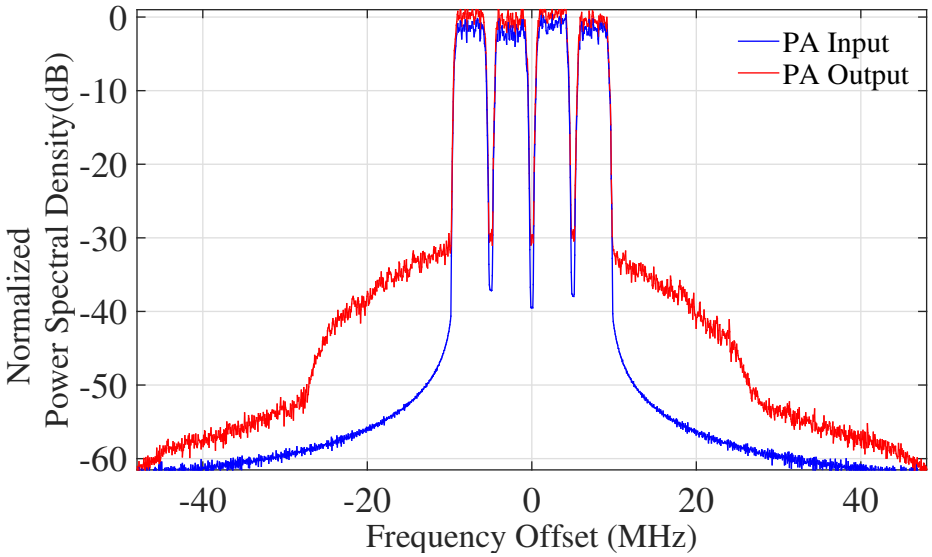


Figure 4.7: PA input and output spectra with a 4-carrier WCDMA signal

Table 4.1: Behavioural Modeling Performance

	Standard LS	**Compressed Training**
Number of Training Samples	8192	681
NMSE(dB)	-45.3	-45.3
Complex Multiplication	26,606,080	2,270,440
Execution Time(second)	11.473	0.092

of standard fixed-point Recursive least squares (RLS) algorithm with consecutive training samples or selected samples, respectively. The results of NMSE versus the number of training samples are shown in Fig. 4.8 for both the traditional and proposed method. The slightly improvement of compressed algorithm using an increased number of training samples is due to averaging of measurement noise. The multiplication operations according to the formula in (4.2) and execution time by a 2.4GHz Core i5 CPU with 8GB DDR to calculate the coefficients are shown in Table. 4.1. The number of training samples is reduced by ratio 12 from 8192 to 681 samples without accuracy degradation. Thus, the number of complex multiplication operations are proportionally reduced by 12 times. Since the LS algorithm is implemented in Matlab where the calculation is performed in serial processing, the execution time is shortened by 124 times (about the square of 12). However, the execution time of a LS algorithm in a practical embedded system will not only determined by the number of multiplications but by many other factors.

$$\sum_{k\in K}\sum_{p\in P}\sum_{q\in Q}\omega_{kpq}x(n-p)|\ x(n-q)\ |^{k} \tag{4.4}$$

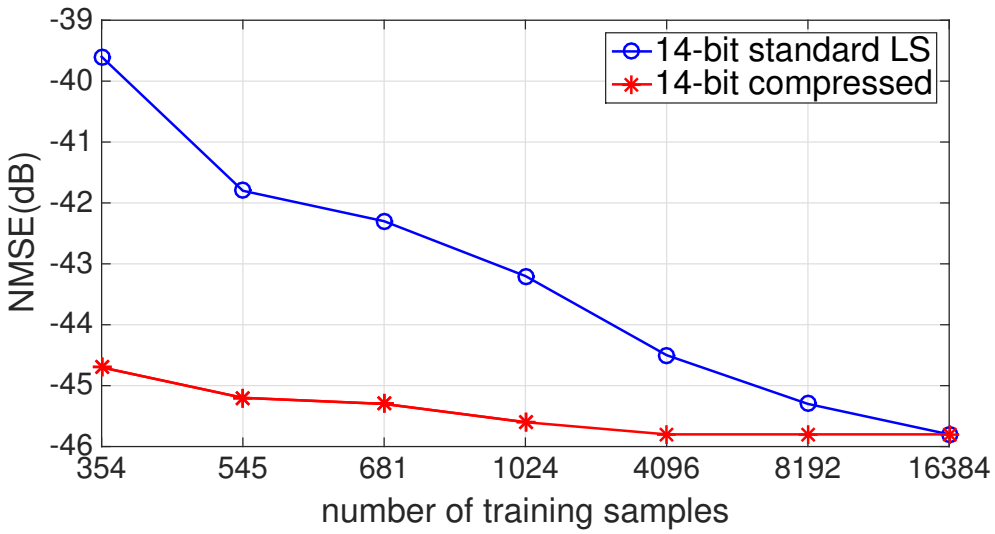


Figure 4.8: Doherty PA Test with 16K complete vector

4.6 Conclusion

Faster, smaller, and cheaper design targets always apply. In future applications of DPD systems for wireless communications, lower power consumption, and faster convergence or coefficient adaption are the challenges we are going to face. The number of complex multiplication operations is the important criteria for evaluating the algorithm complexity. As for the LS-based linear optimization algorithm, the number of complex multiplication operations is related to the number of basis functions and the number of training samples. There has been considerable effort in simplifying the Volterra series approximation by reducing the number of basis functions. However, little attention is paid to reduce the number of training samples which is in direct proportion to the number of multiplication operations in least square based estimation algorithms. As introduced in chapter 3, the responsibility of the training samples is to carry the accurate probability information for accumulating the accurate determined equation. It is low efficient to select consecutive samples as it requires many samples to carry the accurate PDF. The modulated signal is random in time domain but carries stable and predictable probability information for the LS

estimation. Thus, a small amount of training samples can be manually selected which fit the predicted PDF. This chapter proposes the method to identify samples which carry the necessary information to model a nonlinear dynamic system based on the theory in chapter 3. This novel algorithm is highly efficient since there is no additional multiplication operation involved. Moreover, rather than other complexity reduction methods such as LASSO [71] [72] and PCA [73] [66], the proposed method doesn't need to go through the whole dataset or iteratively search the optimized model with less basis functions. The training samples are directly selected based on known probability distributions of transmitted signals. The accuracy of the algorithm is verified using experimentally measured data. The probability distribution of the transmitted signal can be obtained in advanced according to Rayleigh distribution with PAPR as the parameter, and store in the RAM for future usage. However, some signal shaping techniques such as crest factor reduction (CFR) will lead to perturbation of the signal distribution. Therefore, the stability of the proposed method using signal shaping techniques should be investigated in the future. On the other hand, LS square algorithm inherently averages the signal noise. A reduced dataset of training samples also degrades the averaging of measurement noise. Thus, as for a dirty observation path, the necessity of averaging noise limits the minimum number of training samples using the proposed method.

Chapter 5

Undersampling Digital Predistortion: A Theory and Practice

5.1 Introduction

As data rates increase significantly in modern wireless systems, wider bandwidth transceivers will be in high demand. Since conventional DPD systems require multiple times the signal bandwidth in the feedback path, the high-accuracy analog-to-digital converter (ADC) becomes the limiting factor for wideband DPDs.

For the past decade, linearisation of the concurrent multi-band PA has been a topic of interest in software-defined radio (SDR) and DPD research communities. For the traditional multi-band DPD, it is not practical to sample the signals in dual bands as a full band signal, due to the sampling-rate requirements for ADCs. Therefore, advanced multi-dimension behavioural models are proposed

to linearise the multi-band PAs in [74, 75]. However, multiple parallel feedback paths are required to capture the concurrent multiple single bands for multi-dimension behavioural modeling. Recently RF ADCs and carrier relocation techniques are proposed to reduce the number of feedback paths in multi-band DPD systems [76, 77]. Meanwhile, the computational complexity of the DPD estimation algorithm increases dramatically because the number of coefficients for dual-band DPD is much larger than that for single-band DPD [78]. Furthermore, another challenge in the future is the power-consumption of DPD in small cells and massive MIMOs [7]. For low-power transmitters, the power consumed by the DPD unit cannot be ignored. High-speed ADCs occupy a considerable portion of the power consumption of DPD system which is a critical challenge for employing DPD techniques within small cells and massive MIMOs in the future.

Previously in [57], it was illustrated that the mismatch of statistical properties between original data and captured data lead to poor DPD coefficients training processing, and in [79], the authors proposed that given an accurate behavioural model, the accuracy of DPD coefficients extracting using least-square-based algorithms depends on information carried by training samples which is related to the statistical properties of the modulated signal. In [37], a multi-rate identification method is first proposed to directly obtain the coefficient of polynomial-based models using high-rate input and low-rate feedback samples. [39] and [36] also proposed a similar multi-rate modeling method to perform the DPD using low-rate feedback samples and high-rate input samples. Furthermore, based on the probability theory presented in [79], and using the multi-rate identification method of [37], [38] briefly reveals that a generalised DPD comprised of a time-series based model and least-square-based algorithms can be directly performed using aliased feedback samples captured from the low-rate ADC without ad-

ditional calculation cost, which is called undersampling DPD in this chapter. Recently, [40] applied undersampling DPD concept to a compact flexible single-chain multi-band DPD structure which eliminates the intermediate-frequency (IF) stage in the feedback path, and moreover, it also replaces multiple feedback paths that are required in the conventional concurrent multi-band DPD system by a single under-sampled ADC.

In this chapter, we firstly present the complete undersampling DPD theory including defining the information carried by DPD training samples and joint probability information in the nonlinear system with and without memory effects. We then prove that the peak-to-average power ratio (PAPR) and relative bandwidth ratio (ratio between the DPD linearisation bandwidth and the signal bandwidth) of the modulated signal determines the simplified information carried by training samples. The detailed system architecture of the undersampling DPD is also introduced with new comprehensive practical concerns. Finally, both proposed theory and implementation are verified by comparable experimental tests using varying signals captured from different sampling-rates.

5.2 Multi-rate Undersampling DPD Basis

The DPD coefficients are typically estimated from equation (3.3). According to equation (3.9) and (3.10), the DPD coefficients depend on two factors: the probability information $\rho(\phi, y)$ and the value of the feedback samples y. This section will explain why a digital predistortion can be performed using feedback samples collected at any low sample-rate and still obtain accurate coefficients, without additional calculation cost. A multi-rate identification method which was proposed in [37] is introduced in matrix form in 5.2.1. Using this multi-rate identification method, Section 5.2.2 explains how the cyclostationary property of

modulated signal self-reserves the same information from the multi-rate training set comprising of high-rate input samples and low-rate feedback samples.

5.2.1 Multi-rate System Identification

Traditional identification schemes for DPD require that the input samples and feedback samples operate at the same sampling-rate. The multi-rate identification procedure [37] allows us to establish the over-determined equation using high-rate input samples and low-rate feedback samples.

In principle, any valid linear-in-parameter polynomial type of behavioral models can be integrated in the DPD system. For simplicity in this chapter, we will use the generalized memory polynomial (GMP) model basis as below:

$$\phi_n^m = \sum_{l \in L} \sum_{p \in P} x(n-l) |\, x(n-l)\, |^p \tag{5.1}$$

$$m = P \times L$$

where P and L indicate the nonlinear order and the memory depth respectively, and m is the number of the coefficients.

Assume $\{x_n, y_n\}$ represents a time-aligned training sample set with the original high-rate input and high-rate feedback, the over-determined equation for the behavioural modeling can be written as,

$$\phi_n^m \cdot \omega = y(n) \tag{5.2}$$

where m indicates the number of coefficients, n indicates the number of samples.

The model matrix form of $\phi_{n,m}$ can be written as,

$$\phi_{n,m} = \begin{bmatrix} \phi_{0,0}(0) & \phi_{0,1}(0) & \dots & \phi_{0,m-1}(0) \\ \phi_{1,0}(1) & \phi_{1,1}(1) & \dots & \phi_{1,m-1}(1) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{n-1,0}(n-1) & \phi_{n-1,1}(n-1) & \dots & \phi_{n-1,m-1}(n-1) \end{bmatrix} \tag{5.3}$$

And the matrix form of y_n is

$$y_n = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{n-1} \end{bmatrix} \tag{5.4}$$

By only selecting the row of the model matrix which is corresponding with the low-rate feedback sample, a multi-rate model $\phi_{nr,m}$ is achieved, where r is the integer undersampling ratio. (Non-integer processing can be achieved by using an interpolation operation). The multi-rate over-determined equation which is high-rate in row and low-rate in column can be written as,

$$\phi_{nr,m} \cdot \omega = y_{nr} \tag{5.5}$$

The matrix form of the multi-rate over-determined equation is shown as (5.6), and for simplicity, an example is shown in Fig. 5.1 where we assume that 2-tap memory terms are used in the DPD model.

The coefficients can be estimated using least square (LS) based algorithms due to its linear-in-parameter nature (5.7). Since there is no additional computa-

$$\begin{bmatrix} \phi_{0,0}(0) & \phi_{0,1}(0) & \dots & \phi_{0,m-1}(0) \\ \phi_{r,0}(r) & \phi_{r,1}(r) & \dots & \phi_{r,m-1}(r) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{nr-1,0}(nr-1) & \phi_{nr-1,1}(nr-1) & \dots & \phi_{nr-1,m-1}(nr-1) \end{bmatrix} \times \begin{bmatrix} \omega_0 \\ \vdots \\ \omega_{m-1} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_r \\ \vdots \\ y_{nr-1} \end{bmatrix} \tag{5.6}$$

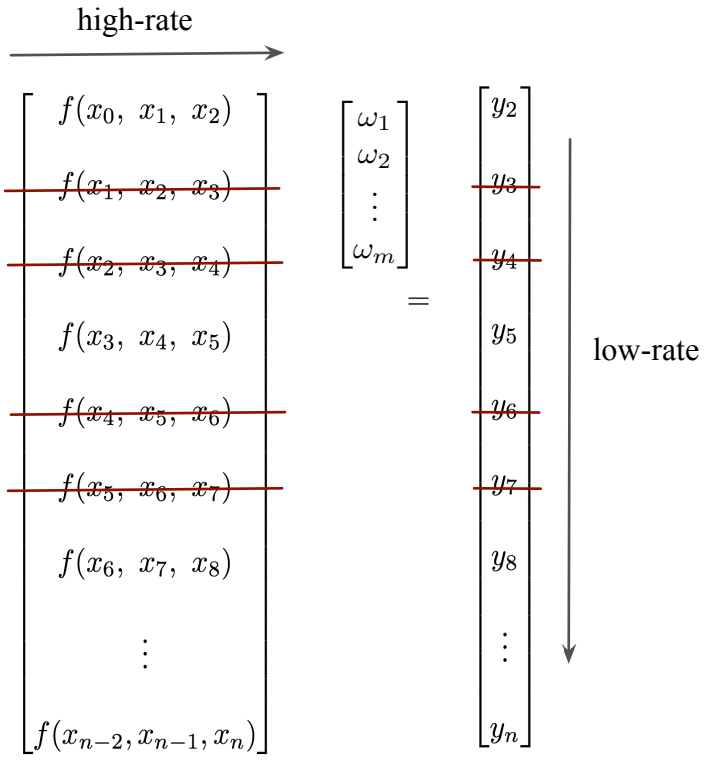


Figure 5.1: An example of generating multi-rate over-determined equation from referenced high-rate over-determined equation

tional penalties for employing this multi-rate modeling method compared to the conventional LS-type (e.g. LS, RLS) DPD parameter estimation procedure, it can be treated as a generalised DPD parameter estimation technique.

$$(\phi_{nr,m}^{\mathrm{T}}\ \phi_{nr,m}) \cdot \omega = \phi_{nr,m}^{\mathrm{T}} y_{nr} \tag{5.7}$$

In a typical LS based coefficients estimation approach, the coefficients are derived by minimizing a so-call error function. Typically, the error function is built in a time domain manner. Sometimes using single sample-based error signal doesnt provide satisfactory estimation performance, then we usually chose sum of squared error which is calculated from a number of training samples. In other words, we use LS approach to find its optimum coefficients that minimize the cost function in time domain. Though LS doesnt see any frequency domain properties, such as aliasing effects due to the insufficient sampling-rate, it does see statistical domain properties of the signals, as explained in (3.10) and Fig. 3.3. As long as the captured short period signal has the same or very similar statistical properties as run-time signals, no matter the sampling-rate of the feedback samples, LS based approaches will do a proper estimation job.

5.2.2 Reserving Information by Cyclo-stationarity of Modulated signals

This subsection introduces that the information carried by low-rate training samples can be self-reserved by the cyclo-stationary property of modulated signals. It proves that the low-rate modulated samples can carry the accurate probability information.

The simplified expansion of the multi-rate joint PDF $\rho(\phi_{nr,m}\ , y_{nr})$ in (5.6) which

includes the high-rate input and low-rate feedback samples can be rewritten as below,

$$\begin{aligned}
&\rho(\phi, y_{nr}) \\
&= \rho(x_{nr}) && \text{(5.8a)} \\
&\cdot \rho(x_{nr-1}|x_{nr}) \cdot \rho(x_{nr-2}|x_{nr}, x_{nr-1}) \\
&\quad ...\rho(x_{nr-p}|x_{nr}, x_{nr-1}, ..., x_{nr-p+1}) && \text{(5.8b)} \\
&p = min\{B-1, m\}
\end{aligned}$$

where r is the integer undersampling ratio, B is the ratio between the linearisation bandwidth and the transmitted signal bandwidth, m is the length of memory taps in the behavioural model.

The term (5.8a) indicates the low-rate nonlinearity term, and (5.8b) indicates high-rate bandwidth terms. It should be noticed that (5.8a) is a low-rate term with respect to under-sampling rate r, and the terms in (5.8b) are still in high-rate. So the joint probability $\rho(\phi, y_{nr})$ in (5.8) represents the information carried by training samples pair, high-rate input samples and low-rate feedback samples.

As well known, a periodic sinusoidal signal with known frequency can be recovered by capturing at any sampling-rate once the frequency is not integer times the sampling rate. It is because of the periodic repeat of magnitude and no disorder after time reordering or aliasing (due to no bandwidth). The multi-rate joint probability $\rho(\phi, y_{nr})$ can be conserved by the property of cyclo-stationarity of a modulated signal in the similar manner. Modern communications signals, such as GSM, WCDMA, LTE-advance and beyond, exhibit high-order cyclo-stationarity due to underlying periodicities introduced through coupling sta-

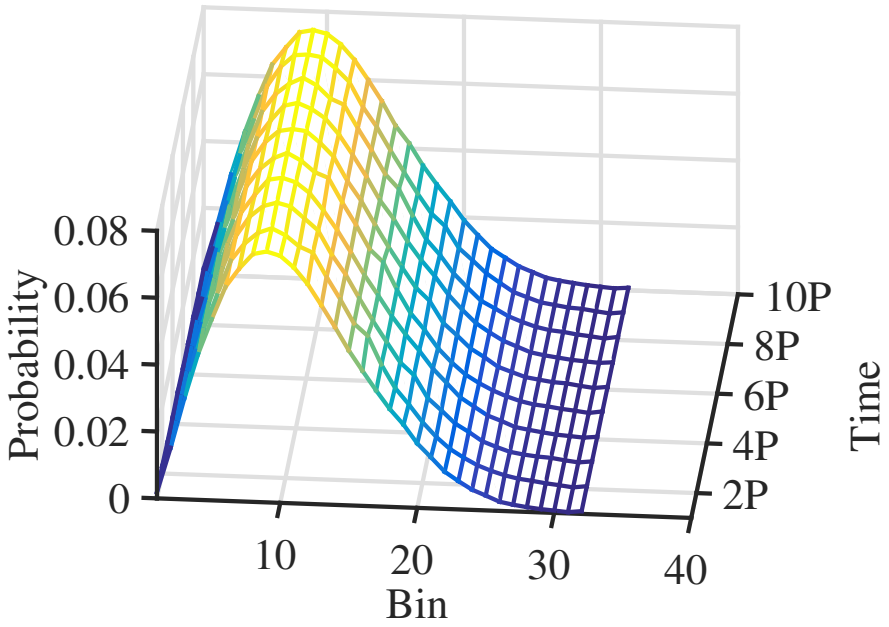


Figure 5.2: The Probability repetition of one-frame 20 MHz LTE signal. P = 8192

tionary message signals, pilot sequences, spreading codes and repeating preambles [80]. The first property of high-order cyclostationary signals is the repetition of probability. Even though the magnitude of a LTE signal is random, the PDF is periodic. As shown in Fig. 5.2, the statistics distribution function of a one-frame, 208 MHz data-rate, 20 MHz bandwidth LTE signal (QAM-256) is approximately constant, for example over each $P = 8192$ samples.

Another property of cyclostationarity is the stationarity in each period P. The probability information of a stationary signal does not vary with respect to time in terms of the high-order stationary process. That means the probability distribution of a continuous-time modulated signal satisfying the high-order stationarity, is not a function of time, but actually a function of length of time occupied by the signal. Therefore the equivalent PDF of low-rate consecutive samples will not lose information after reordering (aliasing) caused by undersampling. The accuracy of the PDF of a stationary continuous signal is related

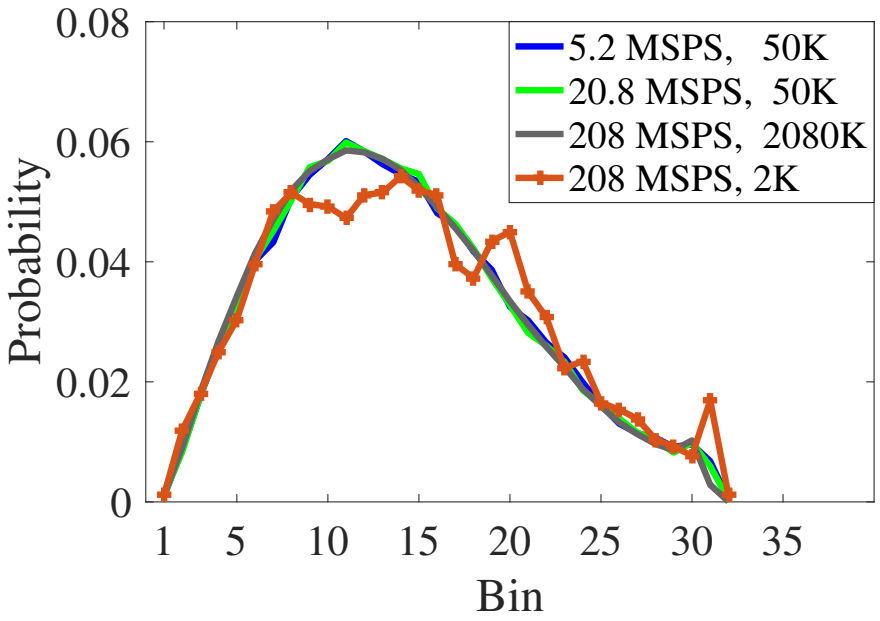


Figure 5.3: Probability distribution of 1-carriers 20 MHz LTE signal $\rho(x_{nr})$ with different sampling-rate and different number of samples

to the length of time occupied by the signal, while for discrete-time signals, the accuracy of PDF is determined by the number of uniform-sampled data. Fig. 5.3 shows the probability distribution of a discrete-time 16-bit 4-carriers LTE signal in different sampling rate and different number of samples. The probability distributions of LTE samples in different sampling-rates are almost identical given that the number of samples is the same. Furthermore, Fig. 5.3 shows that the accuracy of the probability distribution decreases by using less samples even though the sampling-rate stays high. In brief, for a modulated signal satisfying the cyclostationarity, the nonlinearity term (5.8a) is only determined by the number of uniform-sampled data, not the sampling rate.

Similarly, the conditional probability of previous samples and current samples in (5.8b) also satisfies cyclostationarity. Fig. 5.4 shows the conditional probability $\rho(x_{nr-1}|x_{nr})$ given 1/2 of the maximum power of the input signal. As a result, similarly as the nonlinear terms (5.8a), the accuracy of bandwidth terms (5.8b) is also determined by the number of uniform-sampled data, not the sampling

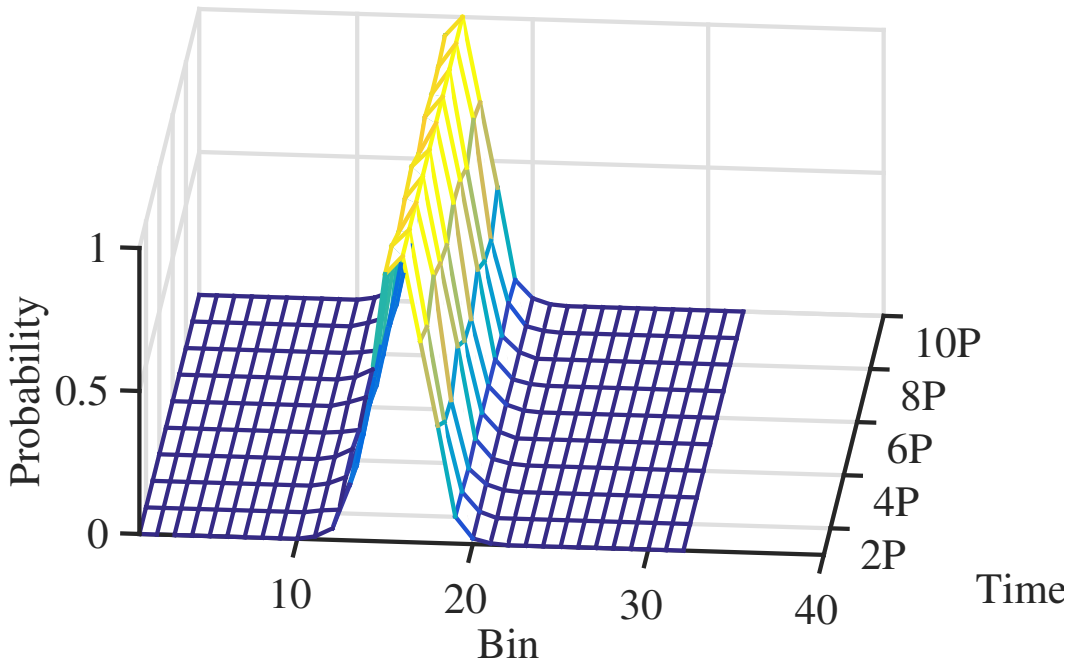


Figure 5.4: The conditional probability $\rho(x_{nr-1}|x_{nr})$ given 1/2 of maximum power of x_{nr}

rate.

To conclude, the information carried by multi-rate training samples (5.8), can be reserved to be the same as the information carried by high-rate training samples.

5.3 Multi-rate Undersampling DPD Implementation

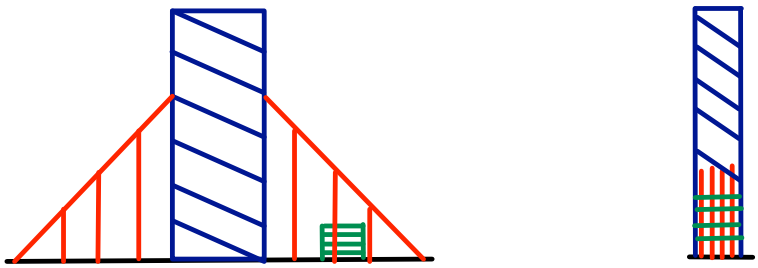
In this section, the implementation of the undersampling DPD with a low-rate aliased feedback ADC will be discussed in terms of both algorithm and hardware. There are some emerging practical challenges for the undersampling-based DPD. These practical concerns include DPD parameters learning structure, hardware structures of the down-convertor, intermediate frequency (IF) selection, band-

width limitation, crest factor reduction (CFR) impact, the number of training samples, robustness with respect to different type of modulated signals.

5.3.1 Undersampling DPD System Architecture

Fig.5.6 illustrates the simplified undersampling DPD system architecture in a direct learning architecture. In order to guarantee the accuracy of the feedback samples captured by ADC in the time domain, the anti-aliasing filter of the ADC has to be the same as the anti-aliasing filter for the traditional DPD which is 3-5 times signal bandwidth. With low-rate ADC, the feedback samples captured by ADC are aliased. However, once the accuracy of feedback samples in time domain and the information of the joint probability information are satisfied, we can say that we obtain a "good" set of feedback samples. Since the feedback samples used in the undersampling DPD are aliased, as shown in Fig.5.5, any interferences could be aliased into the captured feedback samples. Thus, the direct learning structure is preferred for the undersampling DPD because the direct learning is more robust than the indirect learning when noisy measurements are considerable [81]. In the forward data path, a conventional baseband-to-RF transmitter is sufficient. At the feedback path the down-conversion from RF stage to baseband can be done by either using a mixer with a single-ADC (low-IF/high-IF receiver architecture) or by using an IQ demodulator with two ADCs (direct down-conversion receiver architecture). There are some emerging concerns for undersampling DPD that will be discussed in 5.3.2.

A multi-rate time alignment module [40] is required to align the high-rate input samples and the low-rate feedback samples in the time domain. By introducing different phase shift (0 to $r-1$) in the down-sampling processing, where r is the sampling-rate reduction ratio mentioned above, we will have r sets of low-rate



(a) (b)

Figure 5.5: Spectrum of (a) output of power amplifiers with interference and (b) low-rate aliased feedback with interference

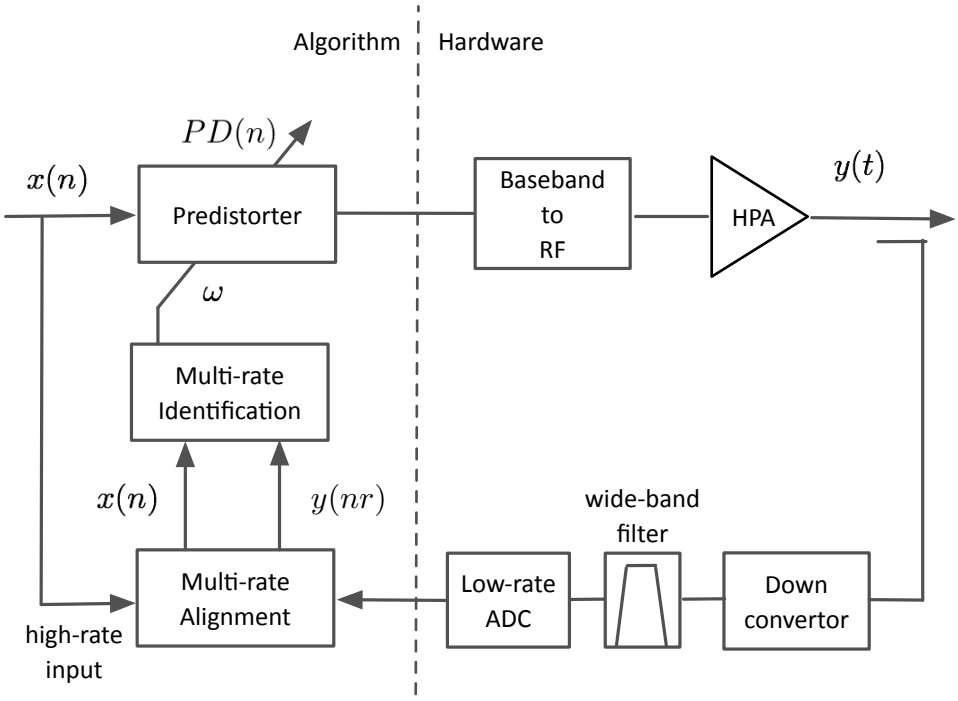


Figure 5.6: The direct learning structure of undersampling DPD

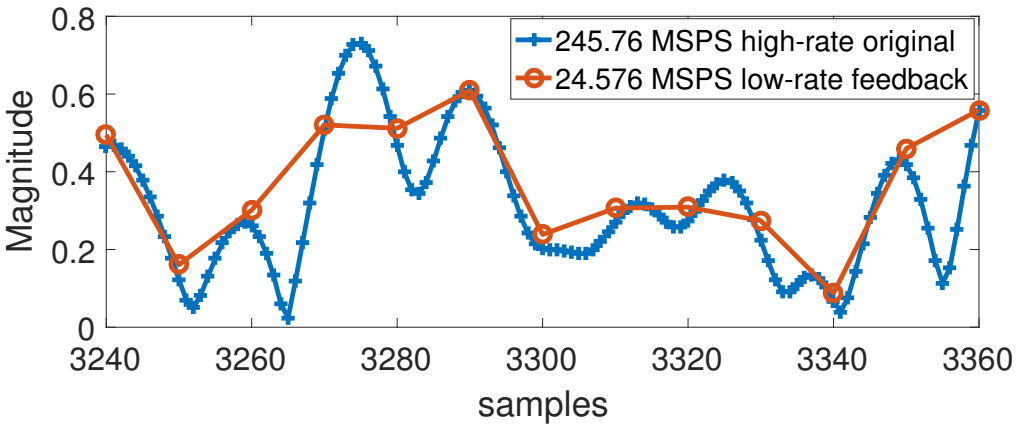


Figure 5.7: Time-aligned multi-rate training samples.

input signals can be utilized to perform cross-correlation-based time alignment. The ith ($i \in [0, r-1]$) correlation result that has the largest peak value indicates the subset $x(i+nr)$ will be used for time alignment. Then the integer delay and fractional delay alignment can be carried out by cross-correction between $x(i+nr)$ and $y(nr)$. The time-aligned high-rate input and low-rate feedback is illustrated in Fig. 5.7. Since the input samples are high-rate, the resolution of this multi-rate time alignment is the same as the resolution of time alignment with high-rate input and high-rate feedback samples. The loop delay between the forward and the feedback paths comes from the DAC, PA, ADC, RF components. In most cases, only the delay from DAC and ADC (mainly from the FIFO) varies if the corresponding devices (for instance FPGA, DAC and ADC) are reset. Thus, the loop delay doesn't need to be calculated in real-time. Even though the multi-rate time alignment is a considerable calculation, it does not need to be repeated and thus is not a limitation for the undersampling DPD or a reason not to significantly reduce the sampling-rate of ADC.

Without loss of generality, the model ϕ can be any time-series based model, such

as generalised memory polynomial(GMP).

$$\phi_{n,m} = \sum_{k\in K}\sum_{p\in P}\sum_{q\in Q} x(n-p)|\ x(n-q)\ |^k \tag{5.9}$$

where m is the number of coefficients.

Then the multi-rate GMP model $\phi_{nr,m}$ used in training procedure is

$$\phi_{nr,m} = \sum_{k\in K}\sum_{p\in P}\sum_{q\in Q} x(nr-p)|\ x(nr-q)\ |^k \tag{5.10}$$

The DPD coefficients can be estimated using (5.7).

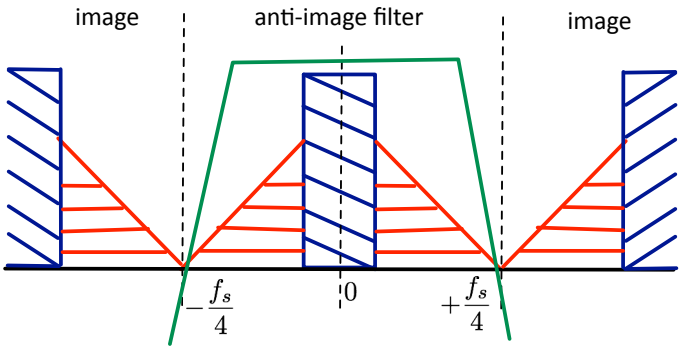
In brief, to implement the undersampling DPD, besides using a low sampling-rate ADC with a wideband filter, the hardware scheme stays the same as the conventional DPD approaches. In terms of the algorithm, a multi-rate time alignment and the multi-rate identification are required. However, since the loop delay is a constant value and can be estimated off-line, the additional calculation cost of the multi-rate time alignment is not an important concern. Meanwhile, the multi-rate identification method doesn't cost additional calculation compared to the traditional high-rate identification schemes.

5.3.2 Practical Concerns of Undersampling DPD

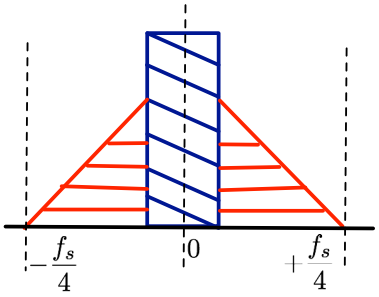
Even though the implementation architecture of the undersampling DPD looks similar to the conventional DPD in terms of both algorithm and hardware, there are some emerging practical concerns on the design of the down-convertor; the IF selection; the minimum sampling-rate and observation bandwidth; the impact of CFR; the number of training samples; robustness with respect to different type of modulated signals.

Regarding the down-conversion from the RF stage to baseband in the feedback path, in principle either real-sampling scheme or complex-sampling scheme can be employed. A real-sampling scheme that comprises of a mixer and single-ADC, demodulating the real samples to IQ samples is achieved in the digital domain which leads to an image in the demodulated IQ samples. A digital anti-aliasing filter which is normally a FIR filter with several taps is used to eliminate the image as shown in Fig. 5.8. However, since the feedback samples in an undersampling DPD scheme are already aliased, if the sampling-rate of the single-ADC is lower than the half of the signal bandwidth, the spectrum of aliased feedback samples will fully fill the Nyquist zone, as shown in Fig. 5.9a. To eliminate the image in this case, the order and the calculation cost of the anti-image filter will be considerably high(the order of FIR filters will be more than 128). In the complex-sampling scheme that comprises of an IQ demodulator and dual-ADC, we need to perform IQ-imbalance correlation. Since, the IQ amplitude/phase imbalance and the dc offset parameters can be calculated using a narrow-bandwidth signal, there is no constraints to implement an undersampling DPD using the complex-sampling scheme running at an arbitrarily low feedback sampling rate.

In traditional DPD approaches, the IF of the down-convertor should be selected such that both undesired spectrum components (including modulation images, spurs, DC offset) and the captured signal are not aliased into the band of interest. For example, the zero-IF in the complex-sampling structure maximises the observation bandwidth comparing to IF-sampling structures, but it requires the additional compensation for the DC offset. In contrast, as the observation bandwidth of the feedback path in the undersampling DPD is not a critical limitation anymore, the undersampling DPD architecture provides more flexibility on IF selection which can lead to the relaxation of some design constraints, for exam-



(a) IQ signal and image from digital demodulation

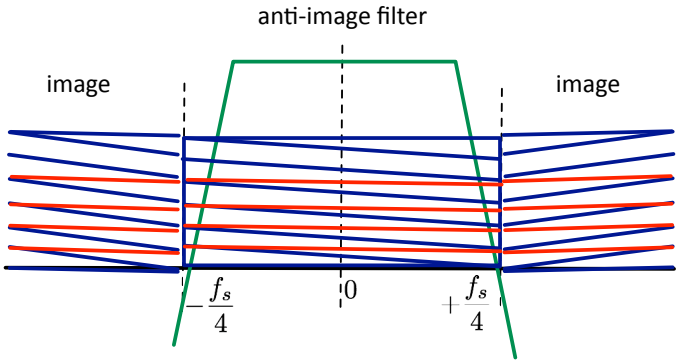


(b) IQ signal after digital anti-image filter

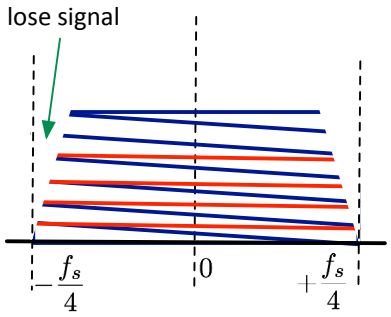
Figure 5.8: Mixer and single ADC running at high sampling-rate

ple using a higher IF to reduce the effect of the DC offset in complex-sampling structures.

Though theoretically an undersampling DPD can be performed using any feedback sampling-rate, the maximum bandwidth and the minimal sampling-rate of feedback path is practically determined by the sample-and-hold (S/H) circuits inherent in the ADC. The feedback path is actually an undersampling system in which the input bandwidth is much higher than the Nyquist bandwidth. To achieve the low conversion-rate, it requires the S/H to hold the signal for a relatively long time with a low droop rate. Droop rate can be reduced by increasing the value of the hold capacitor of S/H, but this will also lengthen the acquisition time which leads to the decrease of the input bandwidth in the sample mode of S/H. So the new design concern of the feedback ADC in the undersampling DPD is to design the S/H circuit to achieve wide input bandwidth with long



(a) Aliased IQ signal and image from digital demodulation



(b) Aliased IQ signal with a low-order anti-image filter

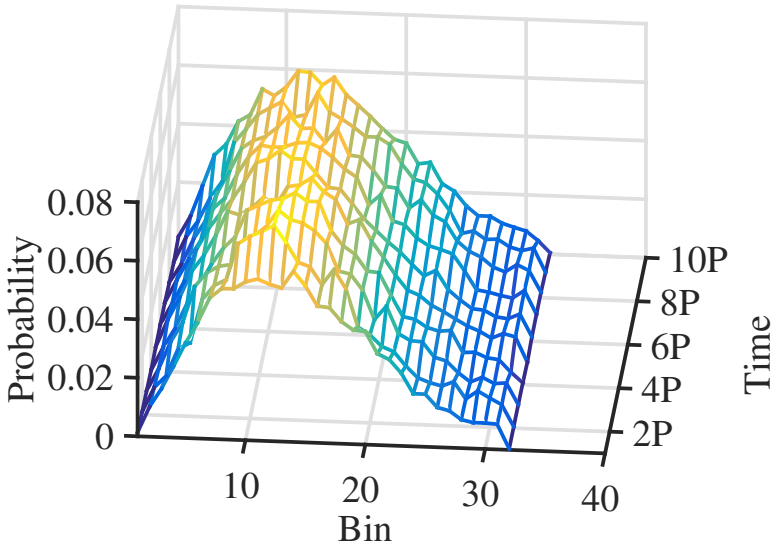
Figure 5.9: Mixer and single ADC running at the sampling-rate lower than half of the signal bandwidth

hold time.

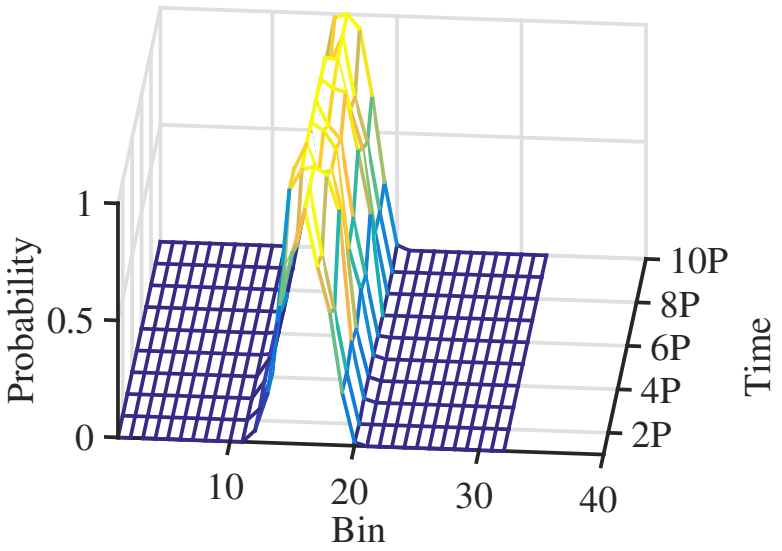
In terms of the algorithm, as introduced in section 5.2.2, the information of modeling can be preserved by the cyclostationary property of idea modulated signal. However, because the high PAPR degrades the efficiency of the power amplifier, the crest factor reduction (CFR) technique is typically employed to precondition the signal to reduce signal peaks without significant signal distortion. However, the CFR will affect the cyclostationary property of ideal modulated signal. As shown in Fig. 5.10, the probability repetition of the signal in Fig. 5.2 is degraded due to 3dB reduction by the CFR. A large peak reduction using CFR could lead to the instability of the probability repetition.

To guarantee the robustness of undersampling DPD, the PDF curve of the low-rate input samples (5.8a) should be verified to be the same as the PDF curve of high-rate input signal. The accuracy of the PDF curve could be improved by employing more training samples. Higher PAPR signal also requires more training samples in DPD training to carry the accurate PDF. As shown in Fig. 5.11, 10K consecutive single-carrier LTE samples with 6.48dB PAPR still maintains the PDF curve of the whole data-set. However, it requires more samples to preserve the probability distribution information of the modulated signal with higher PAPR. Thus, 10K consecutive single-carrier LTE samples with 9.77dB PAPR may represent an inaccurate PDF, as shown in this figure.

Furthermore, in the traditional view of modeling and DPD, higher data-rate brings more hardware cost and computational cost. Fortunately, the higher data-rate also brings more stable probability information. In terms of 4G and beyond, higher data-rate and longer frame period lead to more samples per frame and a more stable statistical property. As proved in section 5.2.2, because of the cyclostationary property of ideal modulated signal, the stability of the



(a) The probability repetition of the LTE signal in Fig. 5.2 with 3dB peak reduction



(b) The conditional probability $\rho(x_{nr-1}|x_{nr})$ given 1/2 of maximum power of x_{nr} with 3dB peak reduction

Figure 5.10: The impact of a 3dB CFR applied to the Cyclostationary of ideal modulated signal

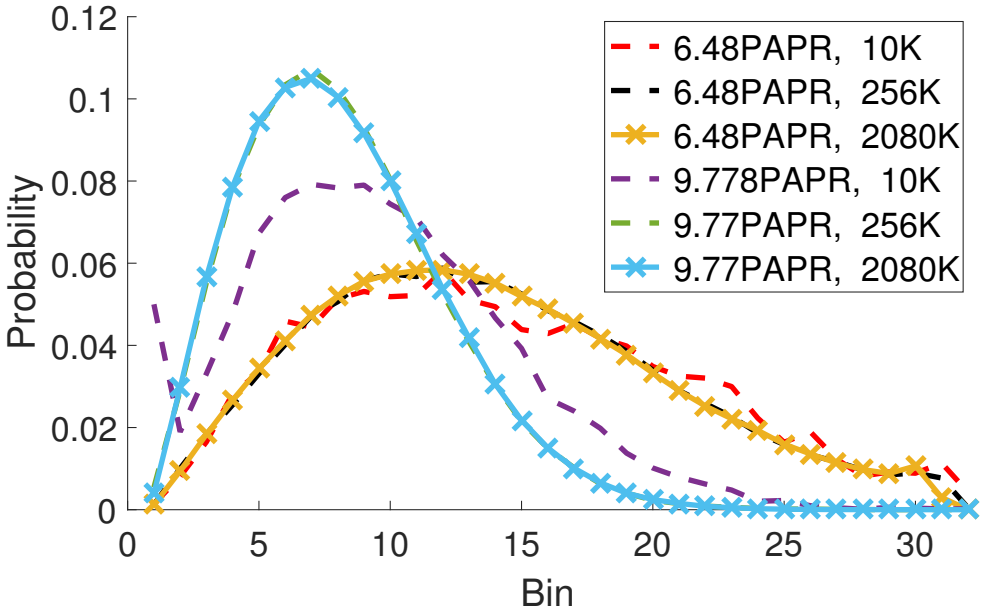
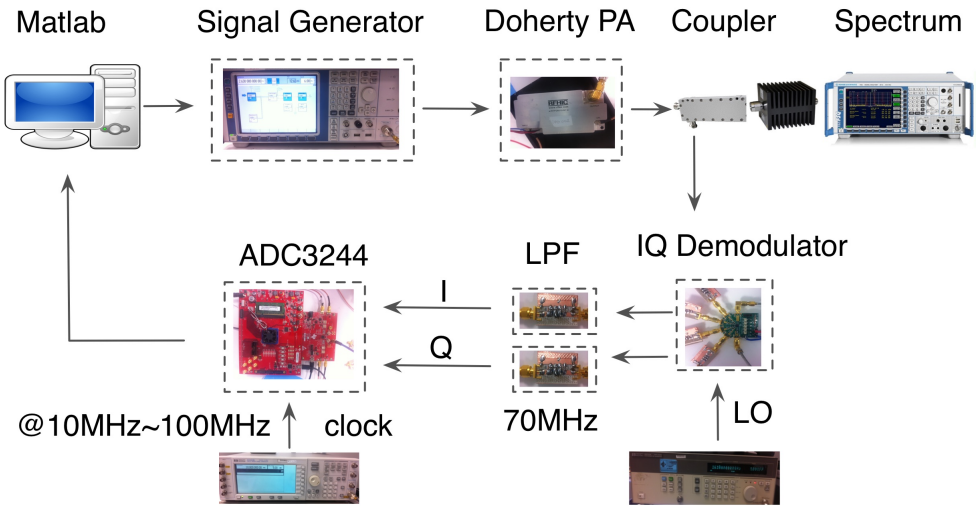
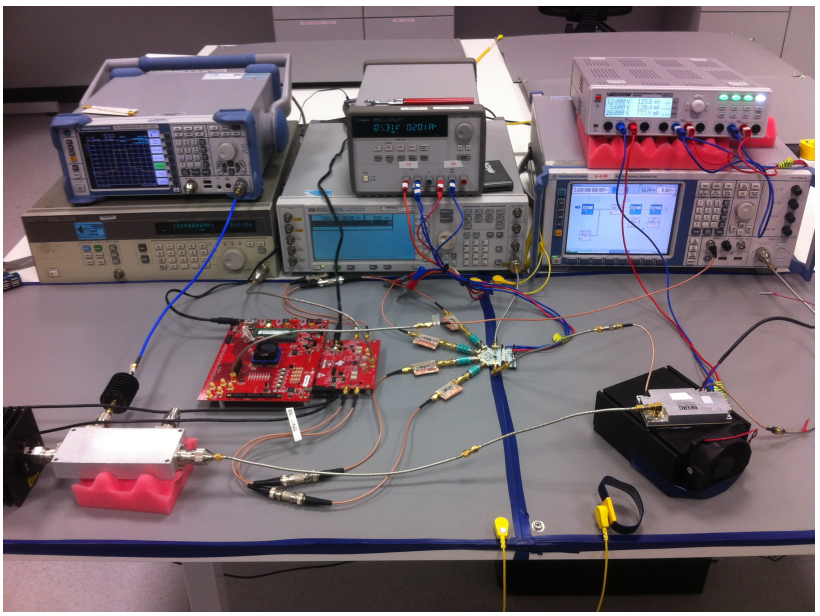


Figure 5.11: The PDF of 1-c 20 MHz LTE signal with different PAPRs and different number of training samples, from 10K to 2080K samples

information of modeling is determined by the number of samples. A GSM signal with single-frame 0.577ms and running at 208 MSPS sampling-rate only contains about 117K samples. Low-rate feedback samples of GSM running at 10.4 MSPS sampling-rate contains only 5.85K samples per frame which may be not sufficient with respect to the fixed-point calculation for wide-band and high PAPR signals. In contrast, a frame of LTE signal running at 307.2 MSPS sampling-rate contains 3072K samples per frame period (10ms). Low-rate feedback samples of LTE running at 6.144 MSPS still contain 61.44K samples per frame period which is sufficient to represent the information of the behavioural model. Thus, in terms of the real-time linearisation application, high data-rate and long-frame modulation signals, such as the LTE, will be more suitable to the undersampling DPD.



(a) Diagram of the test bench.



(b) Photograph of the test bench

Figure 5.12: Setup and photograph of the test bench

5.4 Experimental Result

The undersampling DPD validation test bench was constructed using a Rohde & Schwarz signal generator, a 10-Watt GaN Doherty power amplifier (RTP26010-N1), an IQ demodulator LTC5585, 70 MHz low-pass filters and a dual-input ADC (ADC3244 EVM board), shown in Fig. 5.12. Baseband test signals are generated in Matlab. RF signals centred at 2.6 GHz are sent from the signal generator running at 100 MHz bandwidth to the Doherty power amplifier operating at 40-dBm average power. The output signal of the power amplifier is demodulated and down converted to 20 MHz IF by the IQ demodulator. IF samples (14-bit resolution) are captured by the ADC board as the training sample set, then down-converted to baseband in Matlab. The multi-rate training samples are then obtained after multi-rate alignment. The multi-rate modeling and fixed-point recursive least square (RLS) algorithm in direct learning structure are employed to extract a GMP model [21] with 40 coefficients (order 10 and 4 memory taps). The minimal sampling rate of ADC3244 is 10 MHz, so the feedback samples are captured at the sampling-rate of 100 MHz and 10 MHz.

The first test is to verify the feasibility of the proposed undersampling DPD. A 1-carrier 20 MHz bandwidth LTE signal with 6.48 PAPR value is employed, the PDF of which is also shown in Fig. 5.11. According to Fig. 5.11, 256K feedback samples were captured at 100 MSPS and 10K feedback were samples captured at 10 MSPS (low) and both used to perform the DPD individually. From the test result shown in Fig. 5.13 and Table 5.1, the accuracy of the proposed undersampling DPD stay the same with respect to different sampling-rate. The undersampling DPD with 10 times sampling-rate reduction ratio achieves about -54dB ACPR. There is a small spur at 20 MHz but this was identified as being due to some leakage at the DC offset of feedback path.

Table 5.1: ACPR results using 6.48 PAPR LTE signal which PDF is shown in Fig. 5.11

Number of Training Samples	Sampling Rate	ACPR Left	ACPR Right
128K	100 MSPS (high)	-54.6	-54.2
10K	10 MSPS (low)	-54.3	-53.9

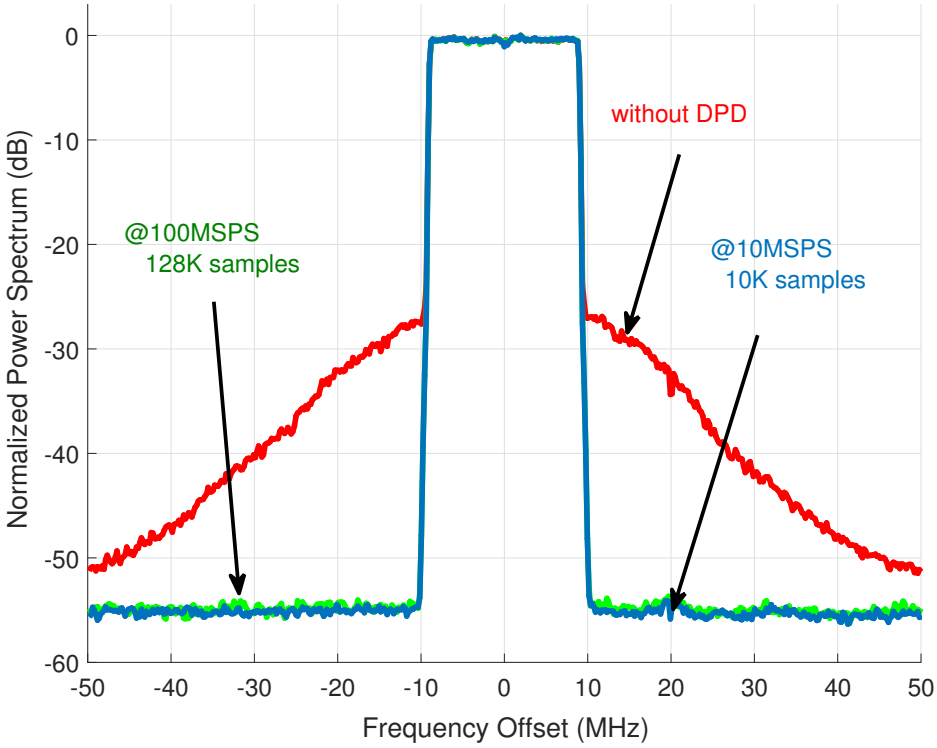


Figure 5.13: Measured spectra of a 47dBm peak-power PA output with 6.48 PAPR LTE signal

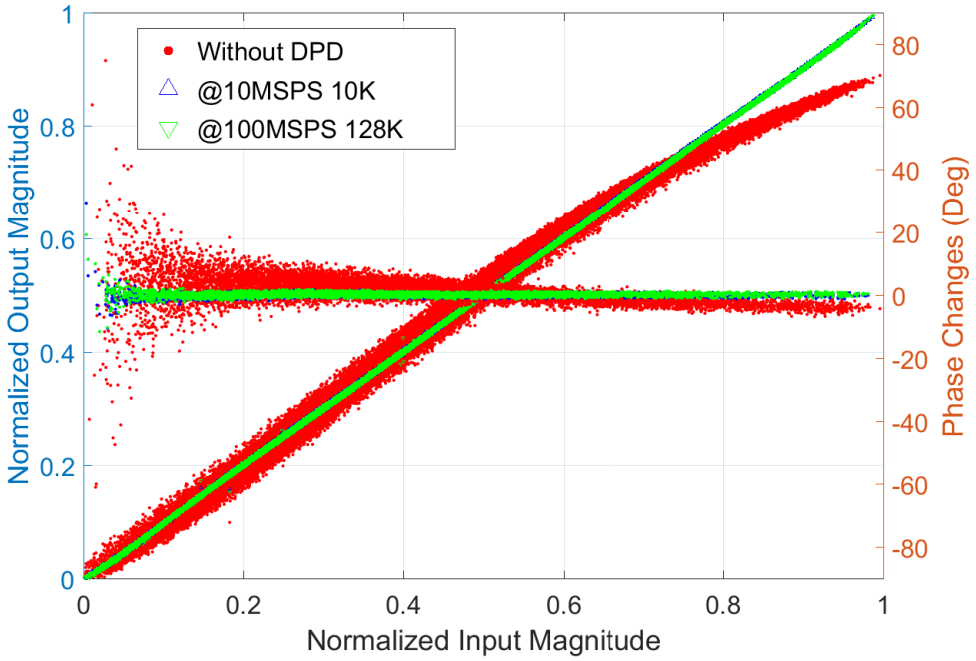


Figure 5.14: AM-AM and AM-PM of PA output with 6.48 PAPR LTE using traditional DPD and undersampling DPD

The second test is to verify the impact of the PDF upon the accuracy of the DPD. According to the analysis in 5.3.2, to guarantee the accuracy of the undersampling DPD, the PDF curve of the low-rate input samples (5.8a) should be verified to be the same as the PDF curve of high-rate input signal. A 1-carrier 20 MHz LTE signal with 9.77 PAPR value is employed in the test which of the PDF is also shown in Fig. 5.11. As before, 128K feedback samples captured at 100 MSPS and 10K feedback samples captured at 10 MSPS(low) are used to perform the DPD respectively. Because the selected 10K samples from the employed 9.77 PAPR LTE signal are not sufficient to carry the accurate PDF information, the ACPR performance of the undersampling DPD using 10K low-rate feedback samples is degraded. The accuracy of the DPD using 128K feedback samples and 10K feedback samples, as shown in Fig. 5.15 and Table 5.2, conforms to the accuracy of the PDF of two feedback sample set shown in Fig. 5.11. The DPD using 128K samples shows better PDF information and better ACLR per-

Table 5.2: ACPR results using 9.77 PAPR LTE signal which PDF is shown in Fig. 5.11

Number of Training Samples	Sampling Rate	ACPR Left	ACPR Right
128K	100MSPS (high)	-53.3	-53.2
10K	10MSPS (low)	-50.0	-49.6

formance. The DPD using 10K samples shows inaccurate PDF information and about 4dB ACLR degradation. Thus, it is important to verify the PDF curve of training samples to determine how many training samples should be used for signals with different PAPR values.

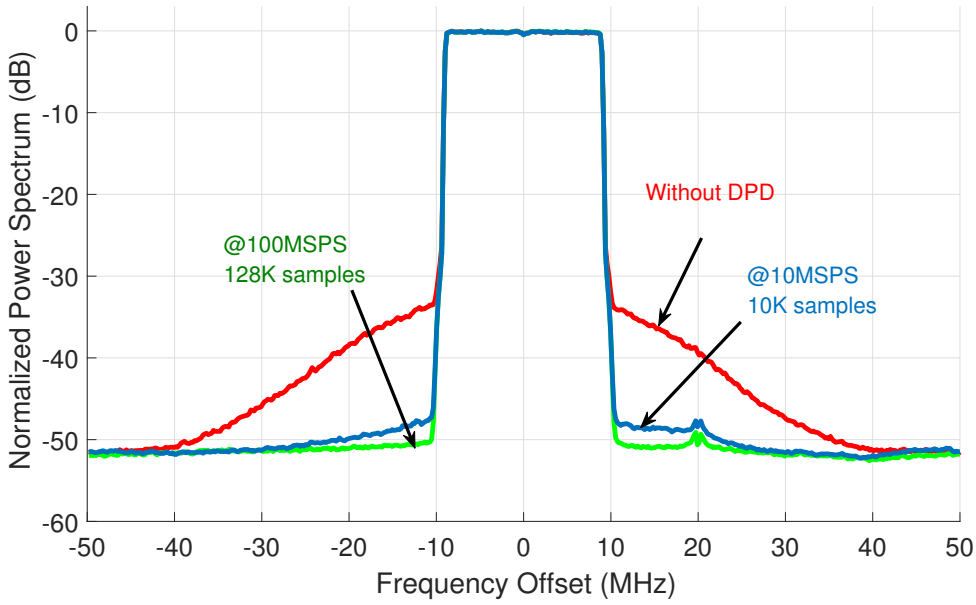


Figure 5.15: Measured spectra of 47dBm peak-power PA outputs with 9.77 PAPR LTE signal

The third test was to verify the theory background introduced in 5.2.2 that the feasibility of undersampling DPD comes from the self-reserving the information using the cyclostationary property of the modulated signal. In this test a 4-carrier 20 MHz GSM signal and a 4-carriers 20 MHz tone signal were used, with a total number of 570K samples for both. The 4-carriers tone signal was selected to have an unstable probability distribution. Although the spectrum of

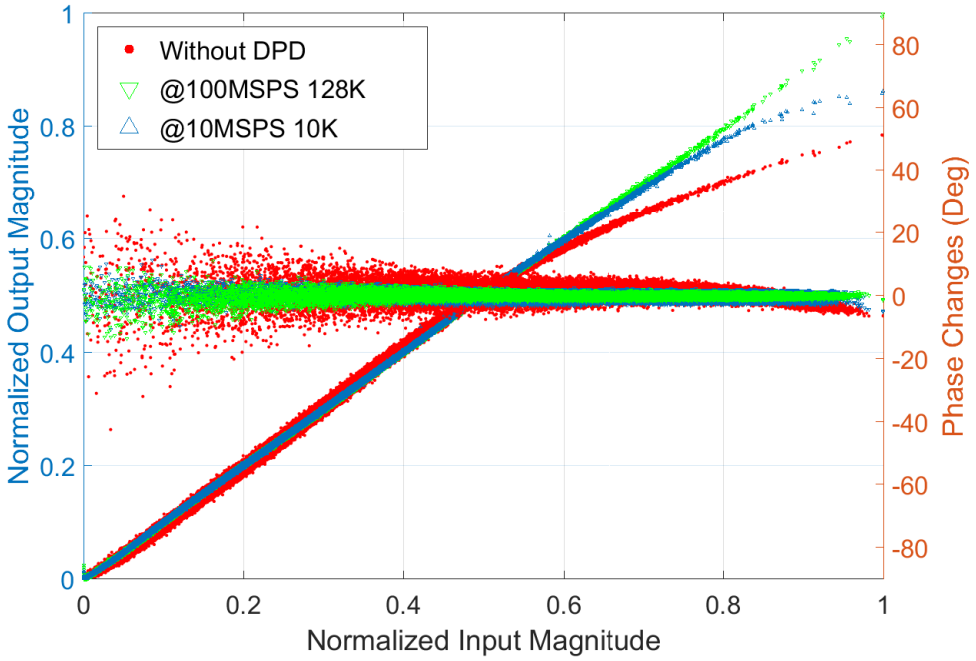


Figure 5.16: AM-AM and AM-PM of PA output with 9.77 PAPR LTE signal using traditional DPD and undersampling DPD

the 4-carrier GSM and the 4-carrier tone signal looks similar, the GSM signal satisfies the cyclostationarity and the tones signal does not. The PDFs of 4-carrier GSM signal and tones signal with different sampling-rate and different number of samples are shown in Fig. 5.17 and Fig. 5.18. It is clear that the PDF of low-rate GSM signal fits the PDF of high-rate GSM signal. However, since the 4-carriers tone signal does not satisfies the cyclostationarity, the PDF of the low-rate tones signal is significantly different to the PDF of the original tones signal.

The spectrum and AM-AM/AM-PM results of the GSM signal are shown in Fig. 5.19 and Fig. 5.20. The spectrum results of the multi-tone signal is shown in Fig. 5.21. The test result using low-rate GSM and tones signal reveals the same fact as the analysis, i.e., for modulated signal with cyclostationary property, undersampling DPD can perform satisfactory linearisation performance, while for the non-cyclostationary signals undersampling DPD assumptions are not

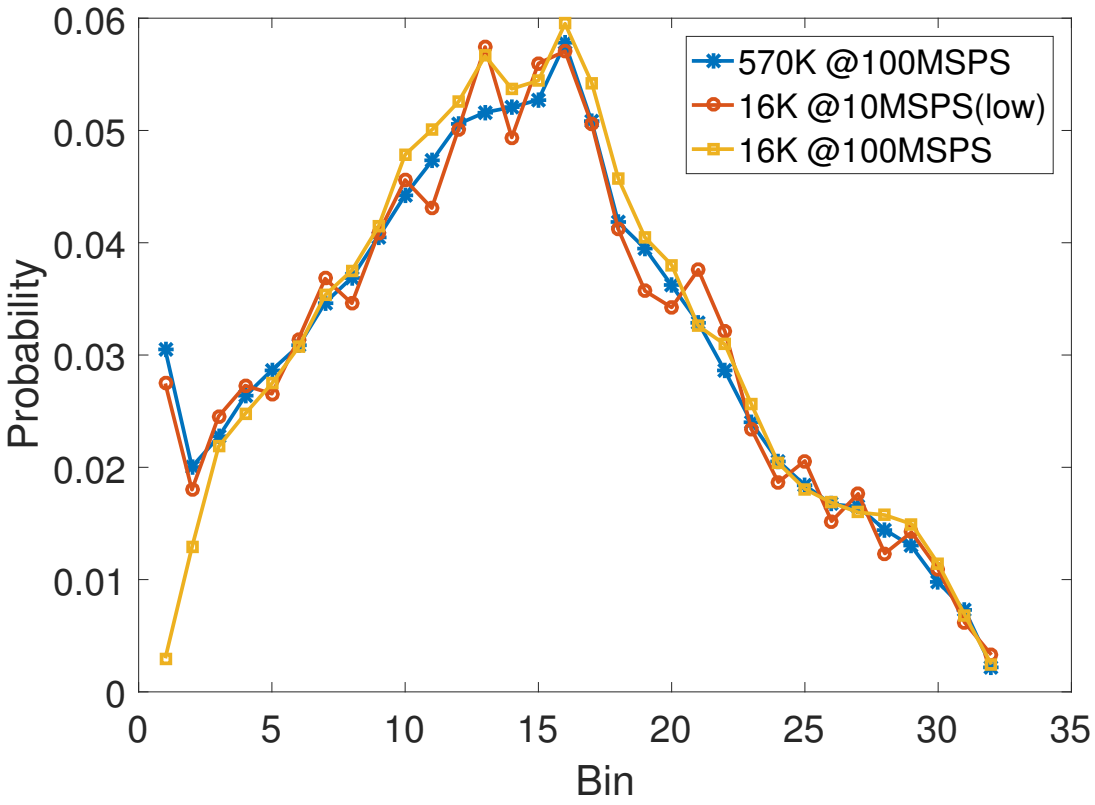


Figure 5.17: PDF of 4-carrier GSM with different number and different sampling-rate

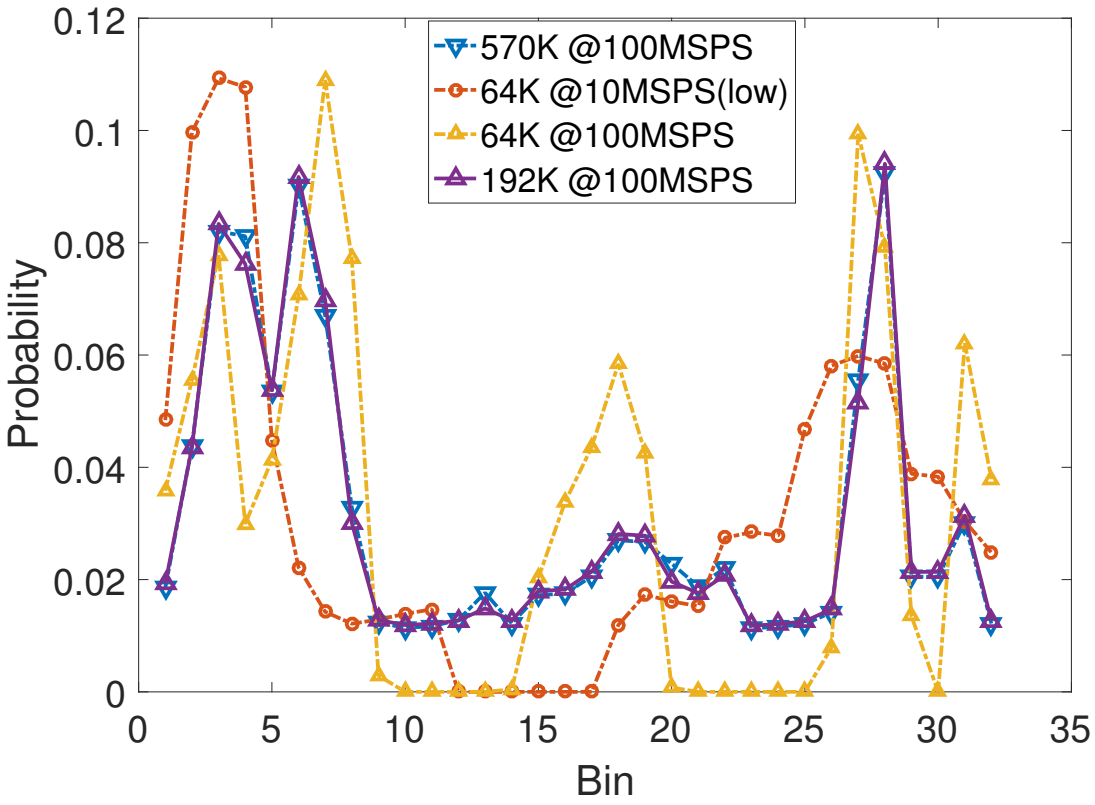


Figure 5.18: PDF of 4-carrier tones with different number and different sampling-rate

Table 5.3: ACPR results using 4-carrier 20MHz GSM and 4-carrier 20MHz Tones signal

GSM	ACPR Left , Right	Tones	ACPR Left , Right
57K 100 MSPS (high)	-53.7 , -53.5	570K (all) 100 MSPS (high)	-54.0 , -51.1
16K 10 MSPS (low)	-53.7 , -53.4	64K 10 MSPS (low)	-42.7 , -40.8

valid leading to non-predictable results. This is not a significant constraint on applicability, as the majority of wireless signals used in the modern wireless mobile networks utilise modulated signals with the cyclostationary property.

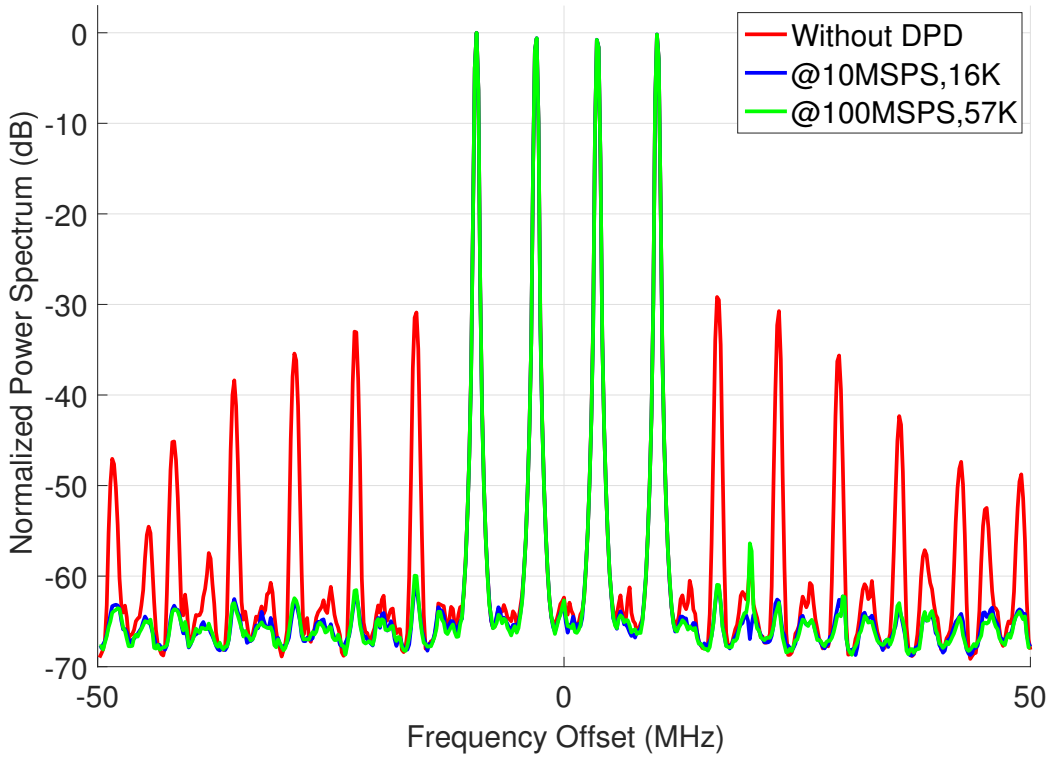


Figure 5.19: Measured spectra of 47dBm peak-power PA outputs with 4-carrier GSM signal

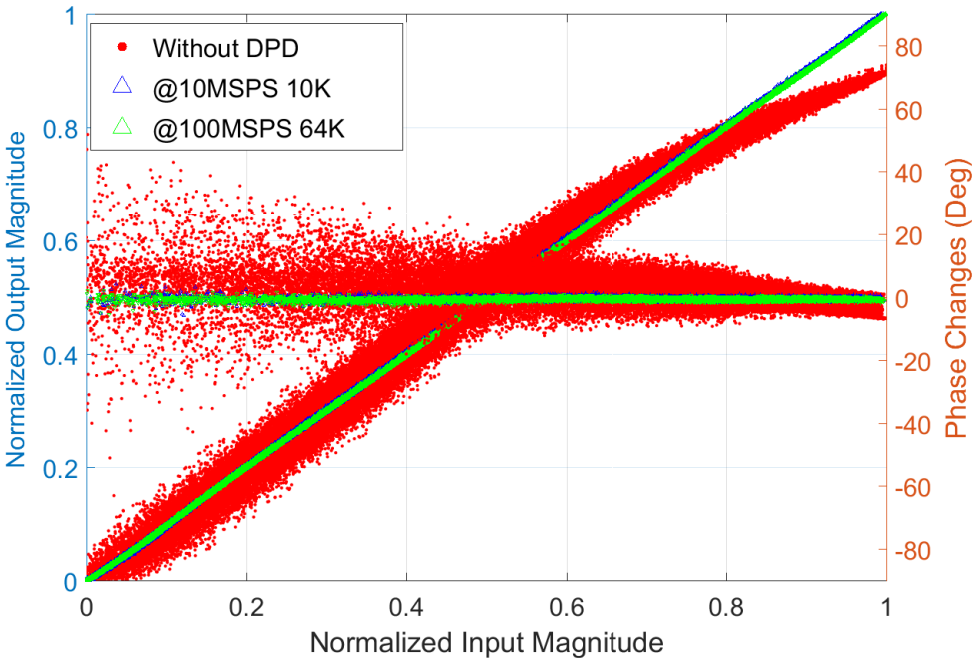


Figure 5.20: AM-AM and AM-PM of PA output with GSM signal using traditional DPD and undersampling DPD

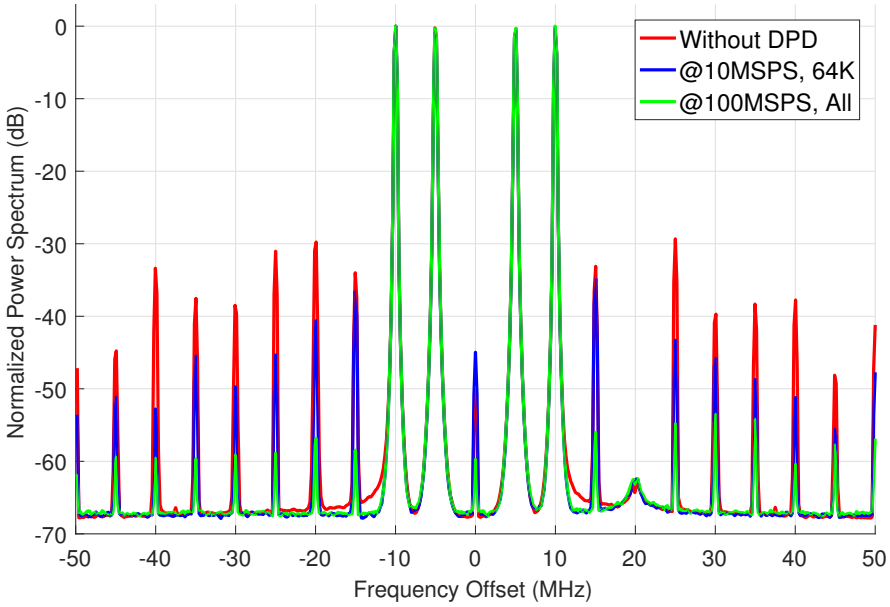


Figure 5.21: Measured spectra of 47dBm peak-power PA outputs with 4-c tones signal

5.5 Discussion

Behavioural model identification is important to a Digital predistortion system. It requires a training sample set comprising of input and feedback samples to estimate the coefficients of the behavioural model. As the training sample set has to represent the nonlinear system accurately, this places challenges on the capture of these feedback samples as data rates expand. Chapter 3 defines the desired information carried by the training samples as a joint probability function of the model and feedback samples. This information is proved to be stable and predictable given the PAPR and the relative bandwidth ratio of modulated signals. Furthermore, the information of multi-rate training samples can be self-reserved to be the same as the information of high-rate training samples by the cyclo-stationarity of modulated signals. It leads to an theoretical and experimental conclusion that the accuracy of the information carried by modulated multi-rate training samples is determined by the number of uniform captured samples, rather than the sampling-rate. These derivations form the theory background of the proposed undersampling DPD.

There is no additional hardware and calculation required in the proposed undersampling DPD because the stationary property of modulated signals takes on the responsibility of reserving the information. So instead of the importance of the sampling-rate, the quality of the probability function of signals becomes the critical factor for the coefficient estimation. It is worth noting that the proposed theory is a universal standpoint for the LS-based modeling method using time-series-based model. Thus, it is also important to verify the quality of the probability function in the traditional DPD training.

Higher data-rate and longer frame period lead to more number of samples per frame and more stable statistical property. So, it is possible to collect enough

training samples in the single frame period of LTE signals with very low-rate ADCs (1 - 10 MSPS). Besides comprehensively practical concerns proposed in this chapter, it is interesting to discuss the minimum sampling-rate and maximum observation bandwidth in a practical DPD system. The minimal sampling-rate is determined by 1) the number of captured samples which determines the accuracy of statistical information and 2) the S/H circuit which determines the accuracy of the sample value. At present, the most critical limit of reducing the sampling-rate in the undersampling DPD comes from the S/H design. The lower droop rate of S/H circuits leading to the longer holding time results in the lower sampling rate of the feedback ADC. On the other hand, the faster acquisition (input bandwidth) of S/H leads to the wider observation bandwidth of the undersampling DPD. A lower sampling-rate ADC will cost less power consumption which is attractive for small cells and MIMO systems. The wider observation bandwidth is attractive for wideband transmissions. Furthermore, since the input bandwidth of S/H circuits can easily reach to several GHz, it is promising to use only one under-sampling ADC with the wideband S/H circuit to replace the entire conventional DPD feedback paths. [40]

More recently, some state of the art works also report a undersampling factor of 20 [48] [41], where similar multi-rate modeling methods as introduced in 5.3 are employed. However, in most of the low-rate DPD work, ADCs are operated in a high sampling-rate, then undersampling DPDs are performed using the digital down-sampled feedback signal. In doing so, some practical issues of employing the undersampling DPD will be concealed. This chapter proposes the theory and implementation of an undersampling DPD system with comprehensive practical concerns on the design of the down-converter (mixer or demodulator); the IF selection; the minimum sampling-rate and capture bandwidth; the impact of CFR; the number of training samples; robustness with respect to different type

of modulated signals.

Chapter 6

Extending the Observation Bandwidth of the Undersampling DPD

6.1 Introduction

As shown in Fig. 6.1, the conventional multi-band transmitter architecture, using digital predistortion, requires multiple forward data paths and multiple feedback paths. In this chapter, we outline and present a compact single-chain multi-band digital predistortion (DPD) solution using only one undersampling ADC and a low-pass filter to replace the conventional multiple DPD feedback paths. We generalized this new architecture as shown in Fig. 6.2. Extensive experimental results validate that the DPD-enabled RF transmitter bandwidth can be significantly increased using multi-rate track-and-hold amplifiers architecture as proposed in this chapter. Experimental tests achieved multiple GHz DPD bandwidth with satisfactory linearization performance via very compact

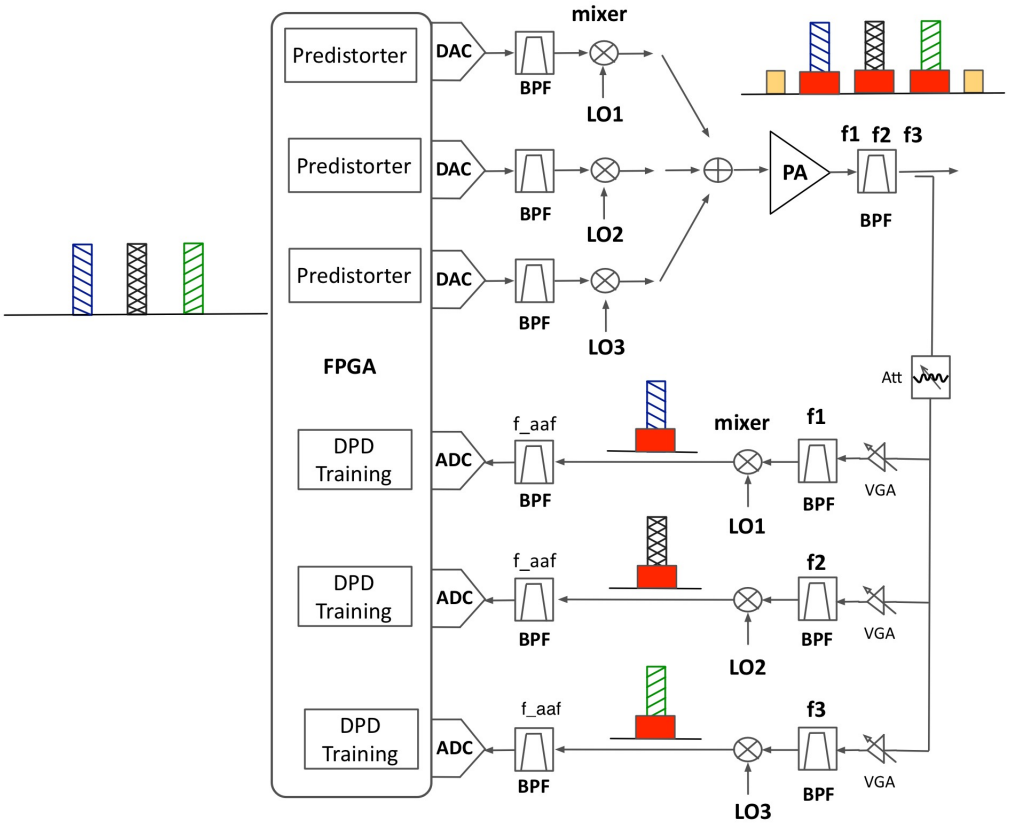


Figure 6.1: Simplified diagram of a conventional DPD-enabled concurrent multi-band wireless transmitter.

sampling receiver that has one undersampled ADC.

6.2 Advanced Compact Single-chain Undersampling DPD System

This section introduces an advanced compact single-chain DPD system. We will address the challenges from hardware and algorithmic perspectives, respectively.

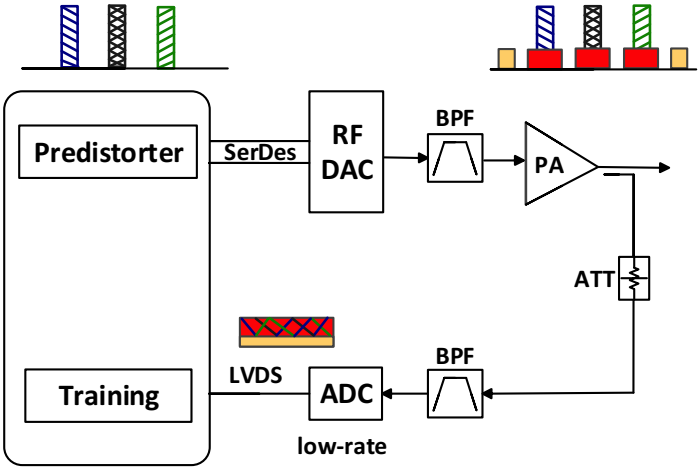


Figure 6.2: Simplified diagram of the proposed single-chain DPD system

6.2.1 System Hardware Architecture Overview

Recall the proposed compact DPD system as shown in Fig. 6.2, we intentionally remove the discrete components based intermediate frequency (IF) stage including mixers, corresponding local oscillators (LOs) and anti-image filters stages in both the forward path and the feedback path. Simply, multiple concurrent baseband signals or continued wideband signals are processed in the digital domain and up-converted to the RF analog domain by RF-DAC. A single-input low conversion-speed undersampling ADC with a bandpass or low-pass filter and the attenuator are employed to replace the entire observation path. To guarantee the accuracy of the analog input of ADC, the bandwidth of the anti-aliasing filter and the sample-and-hold circuits of the ADC stay the same as the interested forward path bandwidth. The sampling rate of ADC fs_{ADC} is fractional times the data forward path RF-DAC sampling rate fs_{DAC}.

The RF sampling ADC has become popular since it provides design flexibility by removing the IF stage. In this proposed architecture, the undersampling ADC is employed to replace the down-converter with the high-speed ADC or a RF sampling ADC. It has the same benefits as RF sampling ADC. Moreover

without consuming such amount of power as RF-ADC, this structure is more power efficient. The hardware cost and PCB size is also significantly reduced.

6.2.2 Extending the Observation Bandwidth by Multi-Rate Track-and-Hold Amplifiers

Theoretically undersampling DPD can be performed using feedback samples collected at any sampling-rate. In practice, the maximum observation bandwidth and the minimal sampling-rate of feedback path are determined by the equivalent capture bandwidth of the track-and-hold amplifier (THA), preceding the ADC which is shown in Fig. 6.3. The THAs function is to sample the input signal at a precise instant and hold the value of the sample constant during the analog-to-digital conversion process. There are two modes of operation in a THA. In the track mode, the voltage on C_H is adjusted to the real-time voltage level at the analog input by charging and discharging the capacitor. During the hold mode, the hold capacitor C_H is disconnected from the input buffer and is expected to retain the voltage present prior to the disconnection. The droop rate, as shown in Fig. 6.4, is the rate that the output voltage is changing due to leakage from the hold capacitor. This combination of acquisition and droop determine the veracity of the captured sample but places opposing constraints on the capacitor C_H. By increasing the value of C_H, it increases the effective hold period but reduces the input bandwidth at the same time. Thus, it is not practical to achieve lower sampling-rate and higher input bandwidth using a single THA.

To extend the analog bandwidth of the ADC, we proposed an undersampling ADC consisting of the multi-rate THAs and a low-rate ADC core as shown in Fig. 6.5. The multi-rate THA circuit is consisted of a high-rate THA1 with

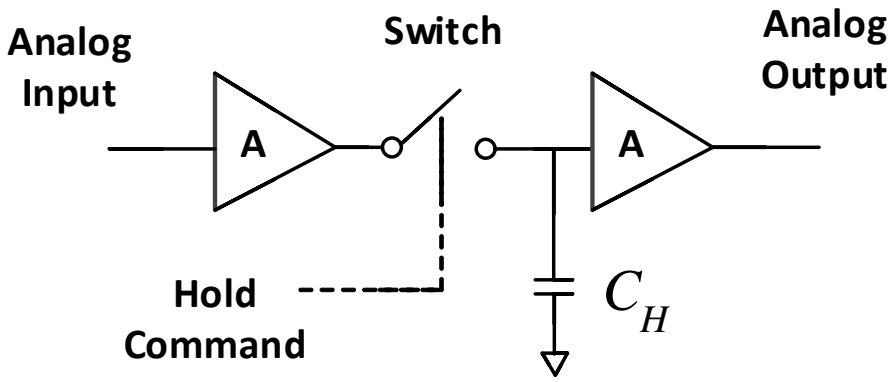


Figure 6.3: Basic open-loop track-and-hold amplifier consisting of a switch, hold capacitor, and input and output buffers.

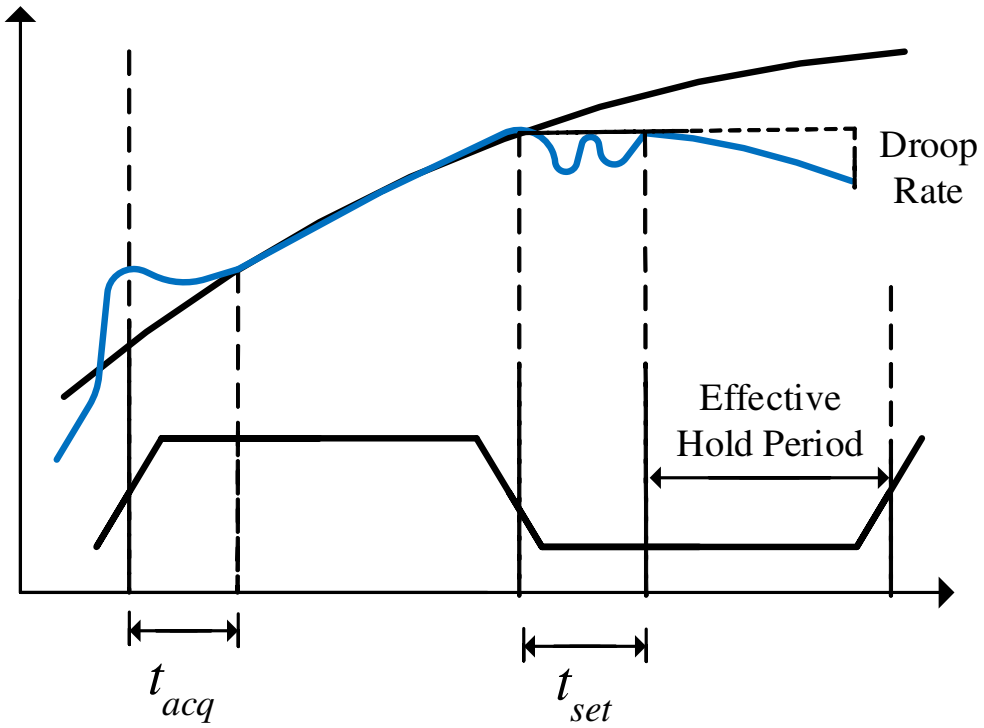


Figure 6.4: Common operation for track-and-hold amplifier

wide input bandwidth and a low-rate THA2 with long hold period. The THA1 holds a captured sample value which is tracked and held by THA2. Fig. 6.6 shows the timing diagram of the multi-rate THAs. The output of the hold mode can be treated as a DC signal. So as for the THA2, it is employed to tracks and holds a short DC signal from the THA1. The long acquisition period from the low-rate THA2 happens at the begin of track mode which will not affect the actual captured sample at the end of track mode of THA2 ($track^2$). The input clock of the undersampling ADC is the high-rate clock for the THA1. It is divided and delayed to a low-rate clock for the THA2 and the ADC core. The sampling-rate of the undersampling ADC is determined by the low-rate THA2. The capturing error of the entire system comes from both the high-rate THA and low-rate THA. The synchronization of the THA1 and THA2 is critical in this design. Due to the variety of propagation delays and digital delays of multi-rate THAs, a delay line is employed to compensate the timing.

6.2.3 DPD Training using Real-number Feedback Samples

Usually in a DPD system, analog demodulation or digital demodulation is required to obtain the in-phase and quadrature (IQ) components at the feedback path. In the proposed compact DPD system, the undersampling ADC directly captures the RF feedback signal. As the IF-stage (analog demodulator) is eliminated, the captured sample value is real number, which introduces two processing challenges.

The first challenge is the different aliasing schemes of complex signals (baseband) and real-number signals (analog). In brief, Fig. 6.7 shows the difference between the aliased baseband complex signal and the aliased analog real-number signal.

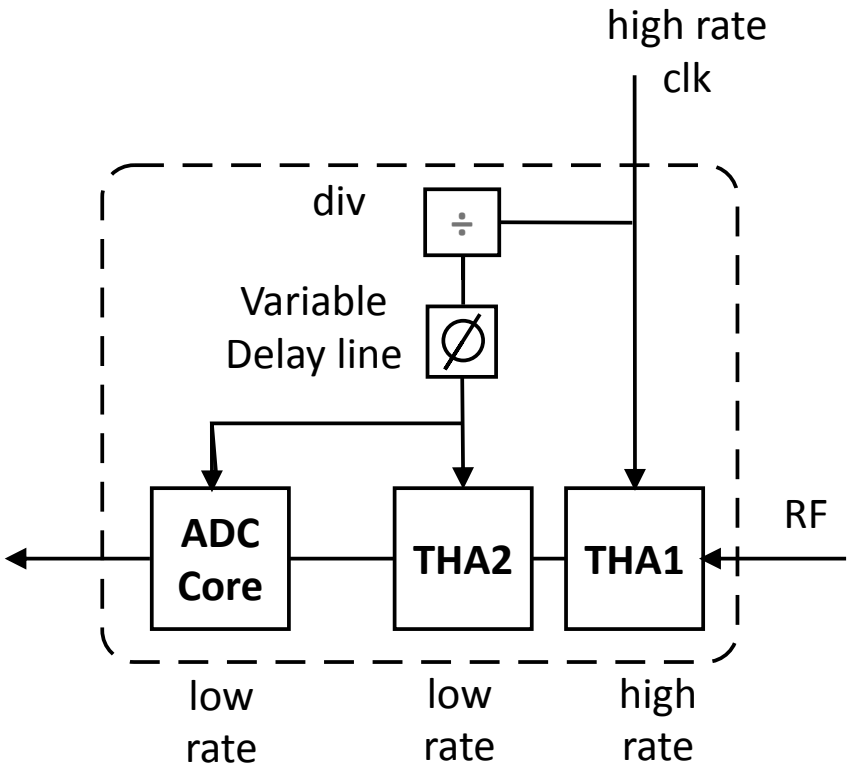


Figure 6.5: Proposed undersampling ADC with Multi-rate THA Circuits.

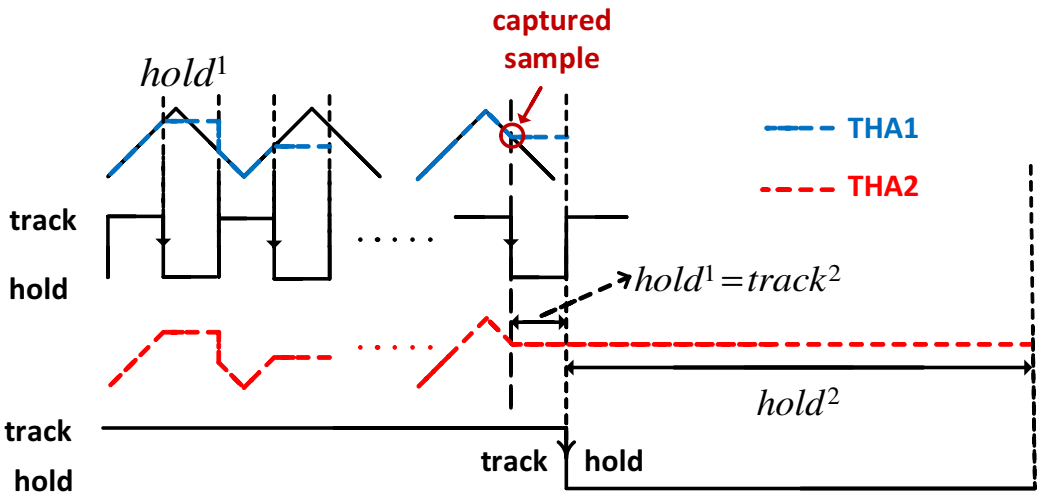


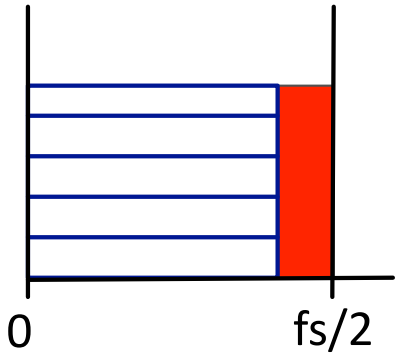
Figure 6.6: Timing diagram of the multi-rate THA circuit

To calculate the coefficients of DPD in the direct learning structure, the errors of the low-rate original samples and the related low-rate feedback samples are necessary, as shown in (3). However, the demodulated IQ samples from the captured low-rate real-number feedback samples are not corresponding to the low-rate original IQ samples, shown in Fig. 6.7.

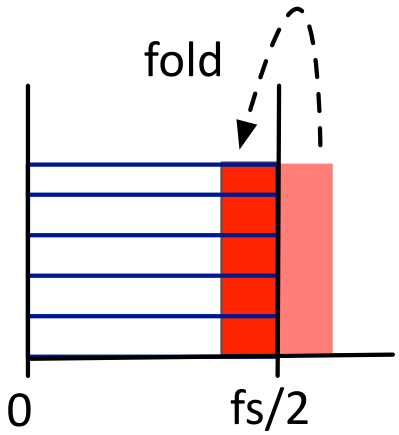
The other challenge using the aliased feedback samples without the analog demodulator is the high orders anti-image filter required in the generalized undersampling DPD. In the undersampling DPD system, if the sampling-rate of the single-ADC is lower than the half of the signal bandwidth, the spectrum of aliased feedback samples will fully fill the Nyquist zone. To eliminate the modulation image in this case, the order of the anti-image filter will be considerably high.

In [82], the authors proved that only the acquisition of either I or Q component of the feedback signal is required for the DPD training. The feedback samples without the phase information(phase of IQ) can be used to estimate the coefficients. So, to address those two challenges mentioned above, the RF signal without demodulation are used to calculate the error function in the proposed DPD training procedure, which avoids two issues mentioned above.

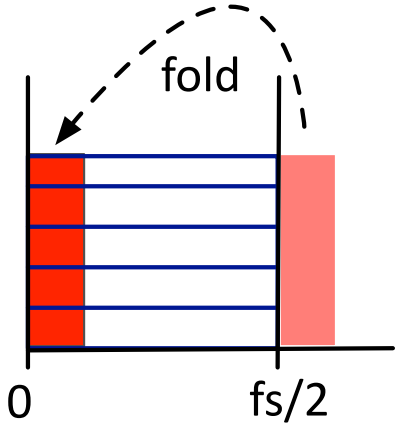
To calculate the real-number error, the original baseband complex samples are firstly modulated to the real-number samples $\bar{x}(n)$. Then a multi-rate time alignment scheme is employed to align the original signal and undersampled feedback signal. Without the loss of generalities, lets assume that the sampling rate of RFDAC is r times the sampling rate of ADC at the feedback path. $\bar{x}(n)$ and $\bar{y}(n)$ represent the real-number high-rate input and feedback samples, respectively. Then the captured feedback samples by the undersampled ADC can be represented as $\bar{y}(nr)$. The original high-rate input signal can be down-



(a)



(b)



(c)

Figure 6.7: (a) high-rate original complex samples. (b) aliased low-rate real-number samples (c) aliased low-rate complex samples.

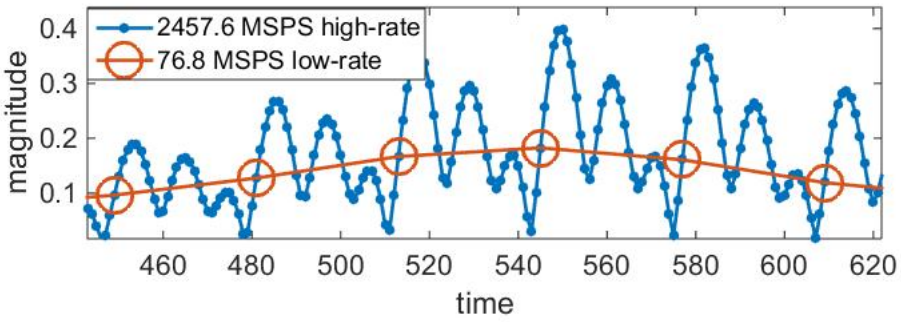


Figure 6.8: An single snapshot example of time-aligned multi-rate training samples.

sampled to produce a low-rate input signal with the same time resolution as the low-rate feedback signal. By introducing different phase shift (0 to $r-1$) in the down-sampling processing, we will have r sets of low-rate input signals that can be utilized to perform cross-correlation-based time alignment. The i^{th} ($i \in [0, r-1]$) correlation result that has the largest peak value indicates that the input down sampled subset x_{in} will be used for time alignment. Then the 2^{nd} time alignment was carried out by cross-correction between $\bar{x}(nr+i)$ and $\bar{y}(nr)$. The time-aligned high-rate input and low-rate feedback is illustrated in Fig. 6.8.

Next the error function used in the direct learning structure is constructed in (12).

$$\bar{e}(nr) = \bar{x}(nr) - \bar{y}(nr) \tag{6.1}$$

where $\bar{x}(nr)$ and $\bar{y}(nr)$ are the real-number samples.

As the delay is calculated by the multi-rate time alignment method, the delay and phase are compensated to the original complex signal to obtain the multi-rate behavioural model ϕ_n^m which means the delay and phase of original signal should be changed rather than that of feedback signal. Then the over-determined

equation using real-number error is obtained in (13) according to [82].

$$M\omega_e = \bar{e}(nr) \tag{6.2}$$

where $M = [(\phi^m_{nr})_I - (\phi^m_{nr})_Q]$, ϕ^m_{nr} is the multi-rate behavioural model building by the high-rate complex input signal $x(n)$, the subscript I indicates the in-phase component and the subscript Q indicates the quadrature component. The coefficients finally can be obtained using the LS algorithm as

$$\omega_e = (M^{\mathrm{T}}M)^{-1}M^{\mathrm{T}}\bar{e}(nr) \tag{6.3}$$

6.3 System Analysis and Practical Concerns

6.3.1 System Analysis

In this section, we will discuss the proposed approach from system perspective including 1) noise performance consideration at the analog chain; 2) power performance consideration at the DPD sampling receiver; and 3) DPD model coefficients estimation complexity.

In a conventional DPD-enabled multi-band RF transmitter, the overall system noise performance of both data forward and feedback paths is typically dominated by the noise introduced by the frequency associated active analog components, like power amplifiers and RF-IF conversion stages. In the forward path, the direct digital-to-RF architecture, particularly RF-DAC devices can be considered as one of the promising technologies to minimize the analog noise by eliminating the discrete components based IF stage. Similar concept can be ap-

plied to the data feedback path, i.e., using direct RF-to-digital architecture, like RF-ADC, at certain larger amount of the financial cost comparing to the utilization of middle range ADCs together with analog down-conversion stage solution (either low-IF or direct down-conversion receiver architecture). The proposed compact approach not only eliminates the IF stage in the forward path naturally by utilizing RF-DAC, but also removes the analog down-conversion stage at the DPD feedback paths by using only one modified undersampling ADC.

Moreover, from power consumption perspective, the current RF-ADC architecture uses multiple converters inside the core, which increases the power consumption and introduces certain interleaving spurs. The undersampling ADC architecture proposed here can achieve multi-GHz bandwidth as well. In contrast, as the undersampling ADC core operates at a low-sampling-rate, the proposed ADC structure consumes considerably less power without interleaving spurs. Utilization of the reduced sampling rate ADC architecture also eases the digital interface design between ADC chipsets and digital host chipsets, like FPGA or DSP. For example, high-speed serial link (SerDes) interface used on RF-ADC will consume considerable high power per channel (150mW per channel running at 3.125Gb/s), and power is also consumed while streaming the received packet into and out of the buffer memory [83]. In contrast, low-voltage differential signaling (LVDS) based digital interfaces which can be used in the proposed undersampling ADC will consume relatively low power per channel due to the significantly reduced amount of data to be transmitted.

The number of DPD model coefficients has been commonly utilized to describe the complexity of DPD model, we follow the same concept here. A straightforward approach for performing multi-band DPD is to build multiple-input and multi-output DPD model to linearize multi-band transmitter simultaneously, and the required number of DPD model coefficients increases significantly

as the number of bands increases. This multi-dimensional expansion of the model coefficients will significantly increase the requirement of digital horse-power for DPD implementations. For example, 50 coefficients for a single-band DPD could grow to several hundred coefficients for a tri-band application [75]. While the proposed DPD architecture is naturally a single input and single output system, which architecturally minimizes the DPD complexity for multi-band linearization applications. In other words, we achieved multi-band DPD capability with considerably much lower-than single-band DPD complexity and power consumption. It is worth noting that, from DPD model perspective, our approach is very friendly and compatible with any linear-in-parameter models, like Generalized Memory Polynomial (GMP) model [21], Dynamic Deviation Reduction based Volterra series models [84], Memory Polynomial model (MP) model, Band-limited model [85] and so on.

6.3.2 Practical Concerns

As we introduced a new design from both hardware and algorithm perspectives, it is worth mentioning some practical concerns for this compact DPD system.

In the multi-rate THA design, the synchronization of the THA1 and THA2 is critical. The relative delay between the clock of THA1 and THA2, which includes the propagation delays and the digital delays, can be calculated or observed by the oscilloscope offline. Nevertheless, a simple way to find the relative delay is to observe the power of captured samples. The multi-rate THA scheme can functionally work when the transition of the track mode to the hold mode of THA2 locates in the hold mode of THA1. Otherwise, the observation bandwidth is degraded. Thus, by observing the power of the captured samples with different relative delay, it is a simple matter to find the relative delay

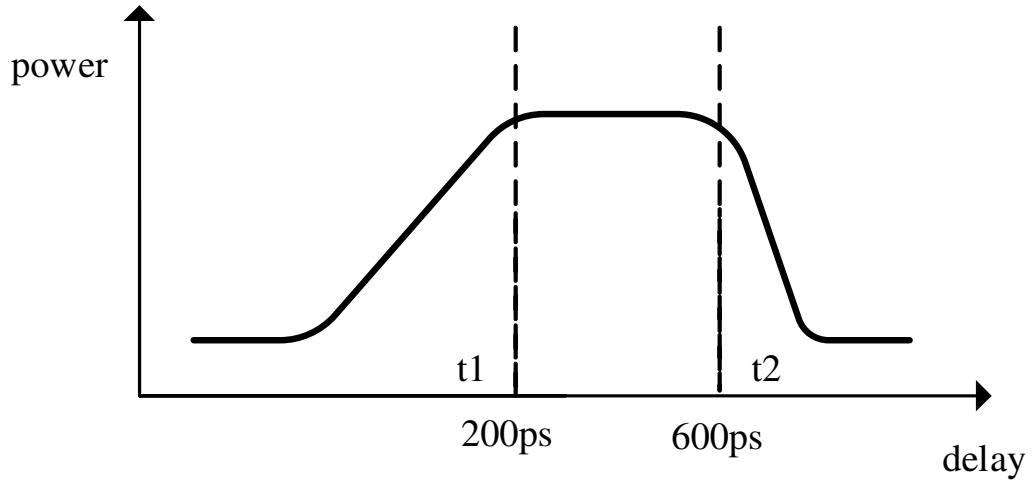


Figure 6.9: The power of captured samples with different delays between the clock of THA1 and THA2

for synchronizing the multi-rate THA. As shown in Fig. 6.9, $t1$ indicates the transition from the track mode to the hold mode in THA1 and $t2$ indicates the transition from the hold mode to the track mode in THA1. The relative delay of the clock of THA1 and THA2 should be within $t1$ to $t2$.

The real-number original and feedback samples are used to calculate the error used in direct learning structure. In traditional DPD system, the phase (IQ rotation) between original signal and captured signal can be calculated as

$$p = (X^{\mathrm{T}}Y)/(X^{\mathrm{T}}X) \tag{6.4}$$

where X and Y are the N by 1 vector of original and feedback complex-value samples respectively. This phase (IQ rotation) value is compensated to the complex-value feedback samples in each iteration. In the proposed system, this phase should be calculated using a narrow-band signal (no aliasing in captured signal) offline and compensate as $1/p$ to the original complex-value original signal in the training process. Then the complex-value original signal can be modulated to the real-number signal. Furthermore, the clocks generated for the RFDAC

and the undersampling ADC should be phase locked to guarantee the constant phase (IQ rotation) between the original signal and feedback signal in each capturing.

The robustness of the proposed system is a practical challenge because any interference and unwanted spurs will be aliased into the feedback samples. The method such as employing the forgetting factor in the recursive least square (RLS) is recommended to improve the robustness of the learning algorithm.

As analyzed in Chapter 5, the performance of the LS-based DPD coefficients extraction approach is largely determined by the probability information of the training samples. Furthermore, as for the modulated signals with cyclo-stationary properties (e.g., 4G LTE signals), the probability information is determined by the number of uniform-sampled data, rather than the sampling rate. Therefore, to properly carry out DPD coefficient estimation, we need a period of the training samples, the probability of which is close to statistical properties of the actual transmitting signal. Fig. 6.10 illustrates an example of the probability function of a tri-band LTE signal, and 16K (16384) low-rate training samples can be selected to extract the DPD coefficients robustly due to the statistical similarity (probability matching between training samples and transmitting samples).

6.4 Experimental Tests and Validation

6.4.1 Experimental Test-bench Setup

To further validate the proposed single-chain undersampling DPD solution, an experimental testbed was assembled as shown in Fig. 6.11. It includes a RF-

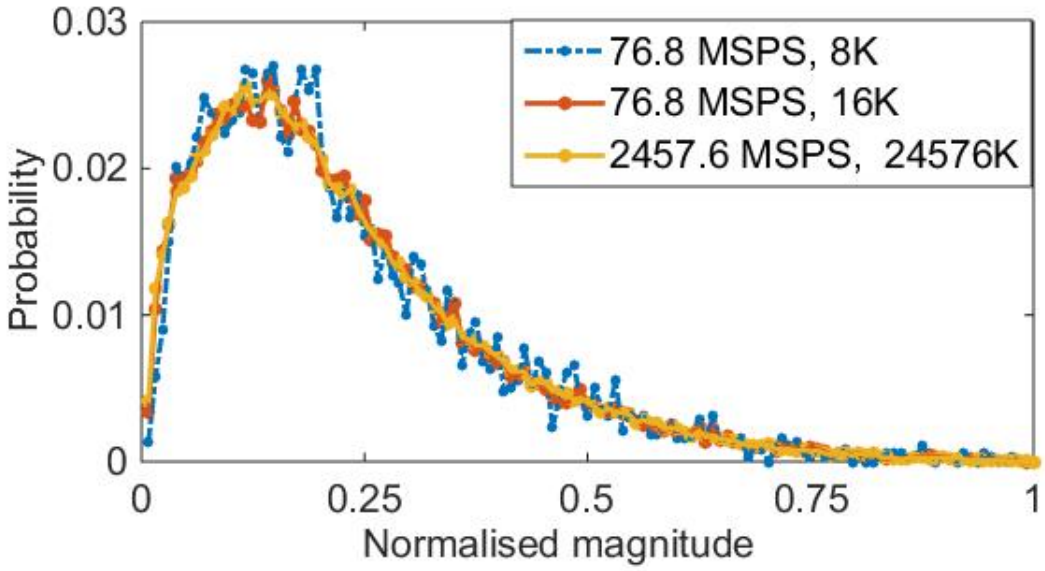


Figure 6.10: Probability distribution with different numbers of training samples

DAC AD9129 (Analog Devices) with a FPGA board ML605 [86], ADC ADS5463 (Texas Instruments) with a TSW1400 data capture board, a pre-driver PA HMC311SC70E, a main PA PAH-1+, a clock generator LMK04808, a wide-band track-and-hold amplifier HMC661LC4B [87], bandpass filters, a low-pass filter and a spectrum analyzer. The RF-DAC was running at 2.4576 GSPS. The ADC was running at 76.8 MSPS without THA1 for the test below 1 GHz. Then it was running at 61.44 MSPS (30.72 MHz Nyquist bandwidth) with THA1 for the test above 1 GHz.

Furthermore, to verify the proposed undersampling ADC with multi-rate THA, the feedback path of the testbed was assembled following Fig. 6.12 for the tests above 1 GHz. The HMC661 ultra-wideband track-and-hold amplifier was employed as the first rank THA (THA1) operating at 1.2288 GHz. This track-and-hold amplifier is claimed to provide maximum 6 GHz track-mode bandwidth with DC-18 GHz observation bandwidth. The second rank THA (THA2) is the THA inherent the ADC (ADS5463). A 1 GHz low-pass filter was placed at the output of the THA1 to optimize the signal-to-noise ratio by reducing the

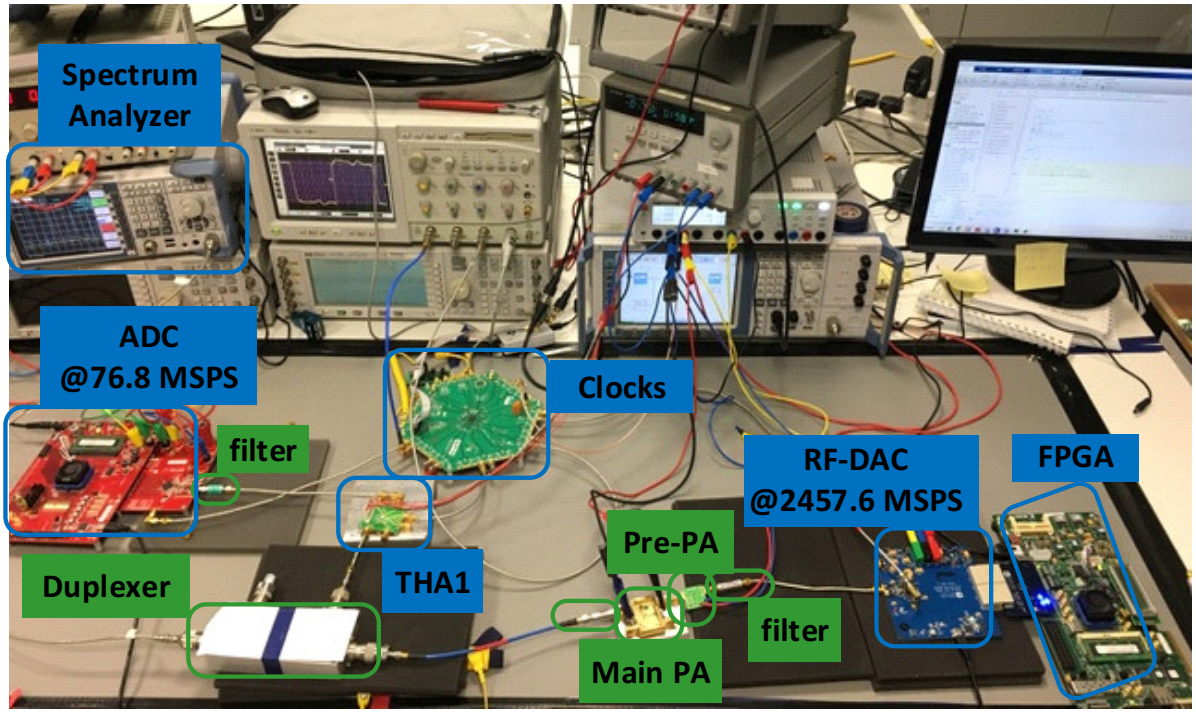


Figure 6.11: Photograph of the test bench

output amplifier noise contribution of THA1 [87]. The behavioural model used for all the tests was the memory polynomial model. The RLS algorithm with 2^{-5} forgetting factor was used for the coefficients extraction. A 10 MHz LTE signal was used in calibration to calculate the phase and delay. During each training procedure, the constant delay and phase value were compensated to the original input samples, rather than feedback samples.

6.4.2 DPD Performance Test Using Tri-band Carrier Aggregated LTE Signals (under 1GHz)

The first validation that we carried our was using tri-band carrier aggregated LTE signals with 11.7 dB peak-to-average power ratio (PAPR). Each of component carrier has 20 MHz instantaneous baseband bandwidth) at center frequencies of 710.4 MHz, 787.2 MHz and 940.8 MHz, respectively. The RFDAC

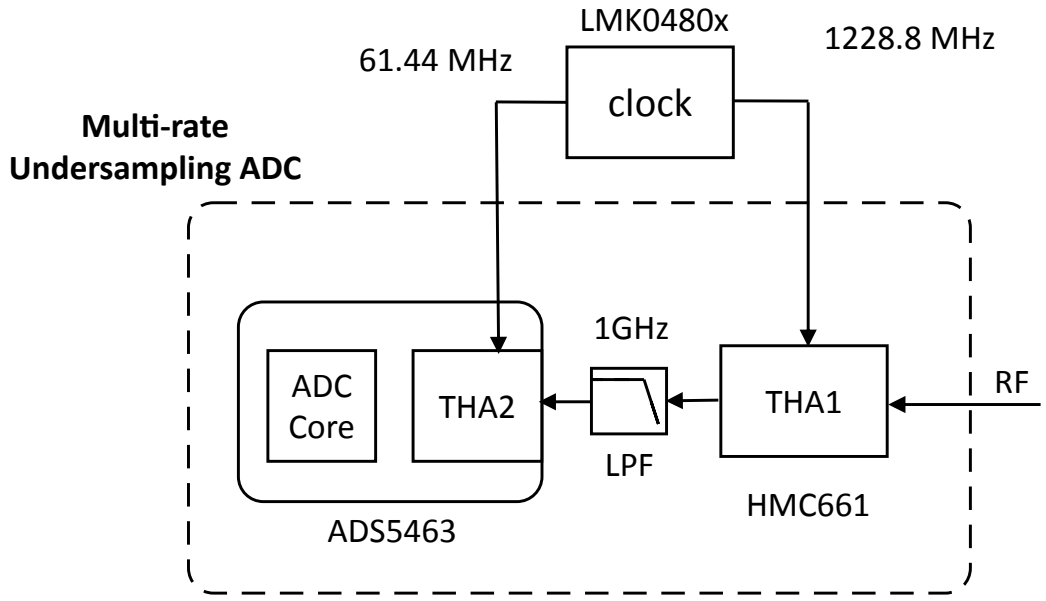


Figure 6.12: The setup of the proposed undersampling ADC structure

was operated in a normal mode to generated RF signal at the first Nyquist zone (0 GHz to 1.2288GHz). The attenuated PA output signal was directly sampled by the ADC with its own in-build 1GHz sample-and-hold circuit. 8K and 16K training samples were utilized to derive DPD coefficients. We implemented a MP model with the nonlinear order = 5 and memory length = 8 (40 coefficients in total). As shown in Fig. 6.13, the PA output spectra with and without proposed compact multi-band DPD approach were captured by the spectrum analyzer. From the testing result, ACPR performance can achieve below -55 dBc [40]. It is worth mentioning that, the probability information of 8K samples of this tri-band LTE signal fluctuates around the accurate probability function of the long data set, shown in Fig. 6.10. The accuracy of DPD confirms to the accuracy of the probability information of the training samples.

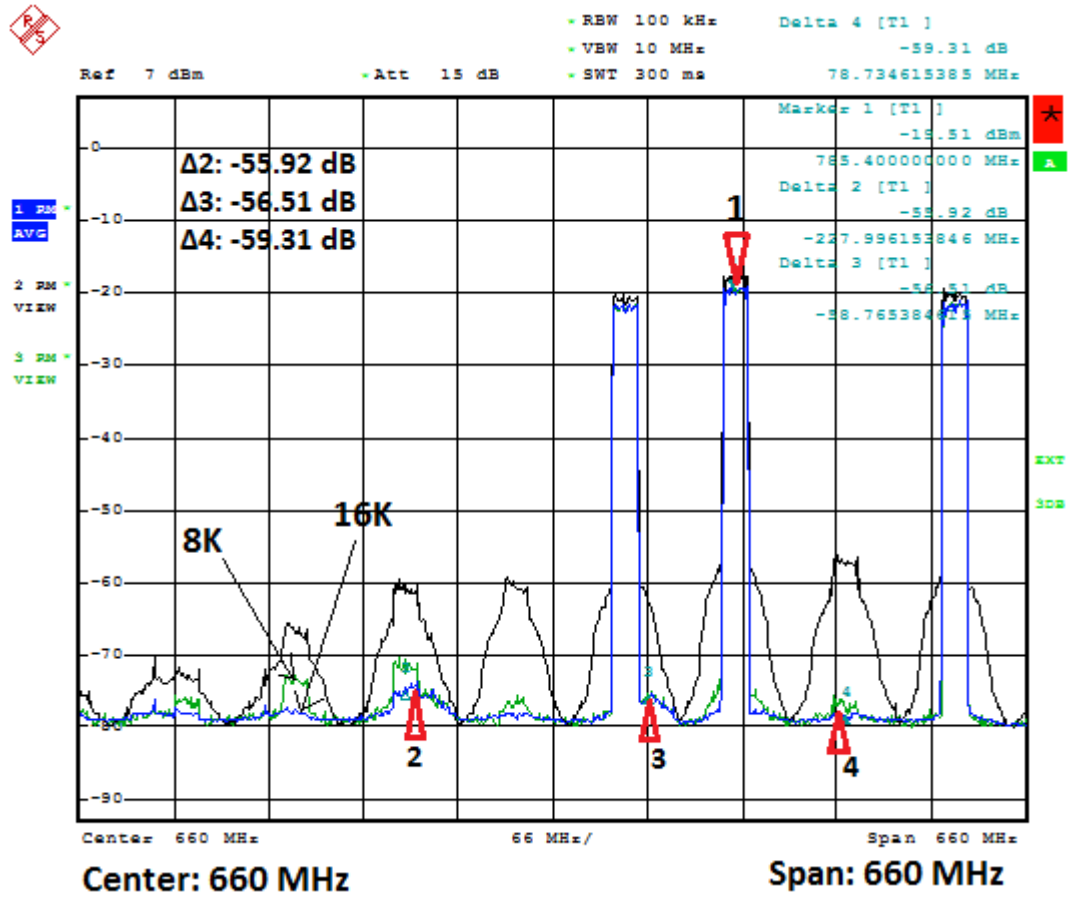


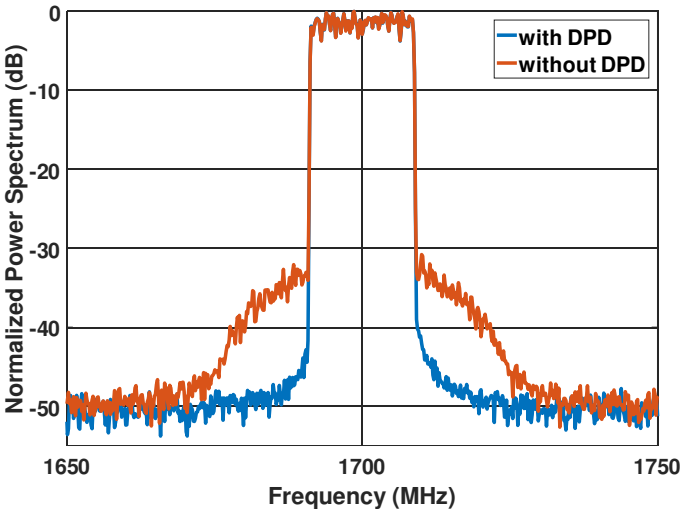
Figure 6.13: Measured spectra of PA outputs before and after proposed DPD using tri-band LTE in 710.4 MHz, 787.2 MHz, 940.8 MHz.

6.4.3 DPD Performance Test Using Tri-band Carrier Aggregated LTE Signals (above 1GHz)

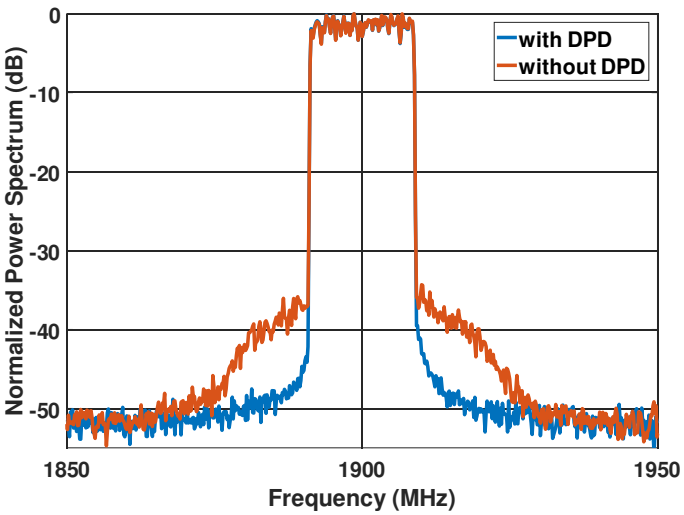
The second test was carried out using tri-band LTE signals(QAM-256) at center frequencies of 1700 MHz, 1900 MHz and 2100 MHz, respectively. We utilized the same MP behaivoral model with 40 coefficients. The RF-DAC was operated in mix-mode to generate the RF signal at the second Nyquist zone (1.2288GHz to 2.4576GHz). However, we found a considerable clock leakage in 2.4576 GHz which folds back to the ADC Nyquist zone at DC in this mix-mode. It decreases the dynamic range of the undersampling ADC due to the unwanted DC injection. The spectrum of PA outputs with and without DPD centering at 1700 MHz, 1900MHz and 2100 MHz are shown in Fig. 6.14a, Fig. 6.14b and Fig. 6.14c, respectively. By utilizing the proposed compact DPD solution, the spectrum regrowth due to the PA nonlinearity and memory effects are eliminated, showing satisfactory multi-band PA linearization performance. Furthermore, we captured small amount of the data to illustrate the AM-AM and AM-PM characterization performance, as shown in Fig. 6.15.

6.4.4 DPD Performance Test Using Continuous Wideband Signal

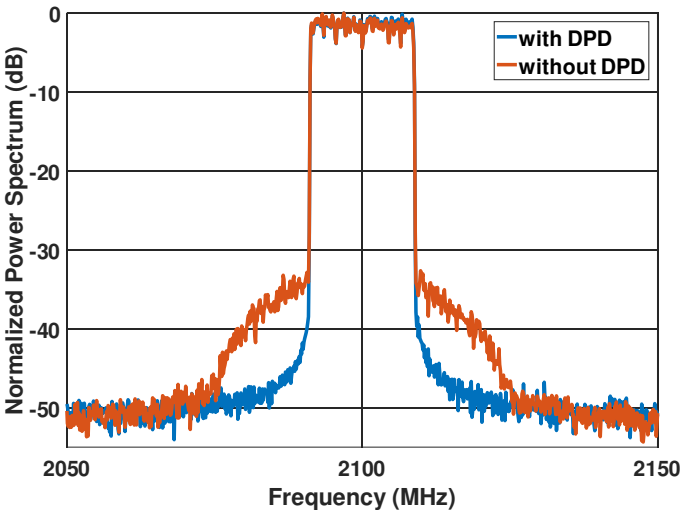
To validate the DPD performance for continuous wideband signal, the third test was carried out using a single-band signal with continuous 100MHz instantaneous RF bandwidth. The centre frequency of the test signal located at 1780 MHz. Preliminary, the ACPR performance was achieved below -47 dB as shown in Fig. 6.16. The AM-AM and AM-PM performance is shown in Fig. 6.17. It is worth re-emphasizing that we only have one low-sampling rate undersampled



(a)



(b)



(c)

Figure 6.14: Spectra of PA outputs before and after DPD using tri-band LTE signals in (a) 1700 MHz (b) 1900 MHz (c) 2100 MHz.

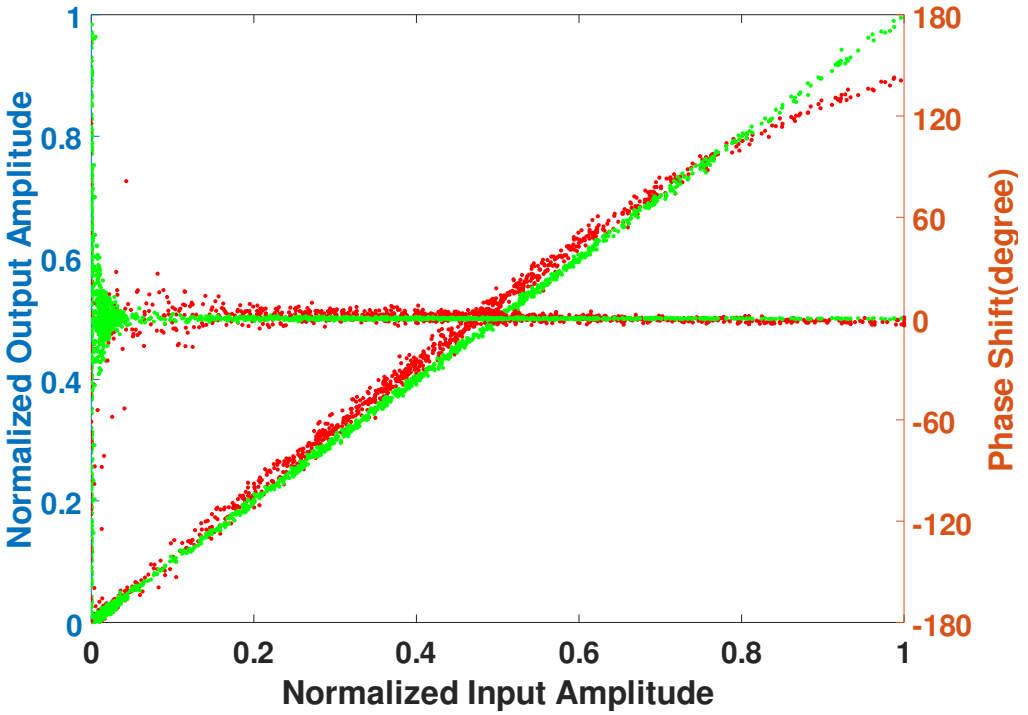


Figure 6.15: AM-AM and AM-PM of PA output of tri-band LTE with and without proposed DPD solution.

ADC (running at 61.44MSPS) and one BPF at the DPD observation receiver path in those tests. At this stage, we were not trying to compete with the best wideband DPD performance in the literatures. We would like to provide a feasibility validation to show that our newly proposed undersampling DPD approach (both HW and SW) can be utilized for wideband DPD application even with heavily aliased feedback signals.

6.4.5 DPD Feedback Path Usable Bandwidth Validation Test

To further validate the capability of the proposed DPD approach regarding the achievable DPD feedback path bandwidth, we re-organized the platform in the following way: utilizing the vector signal generator R&S SMU 200A to replace

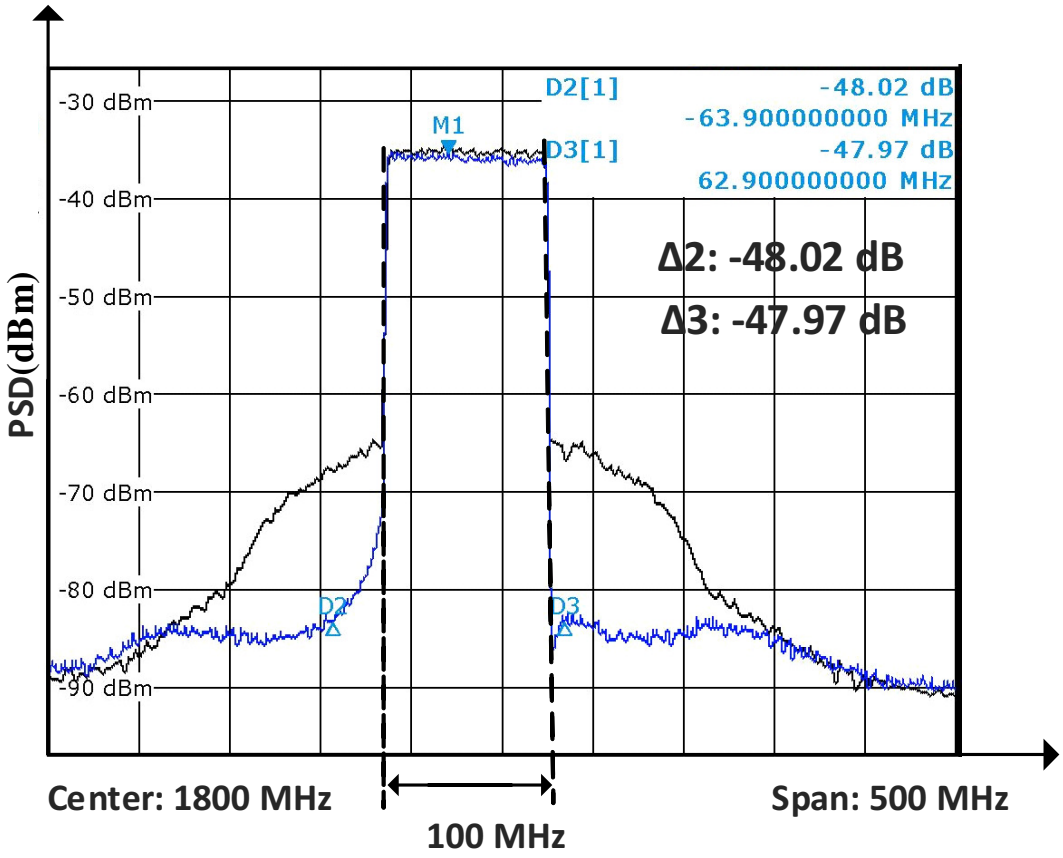


Figure 6.16: Spectra of PA outputs before and after DPD using 100MHz bandwidth signal.

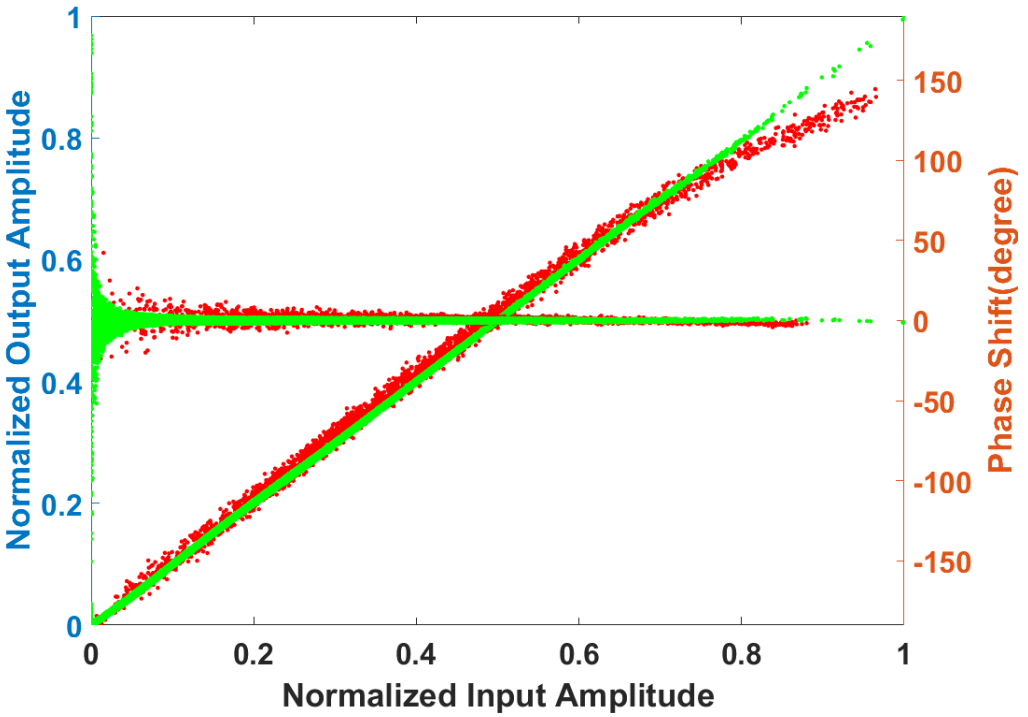


Figure 6.17: AM-AM and AM-PM of PA output of 100 MHz bandwidth signal with and without proposed DPD solution.

the RF-DAC in the forward path and maintaining the single-ADC based DPD feedback path. In this way, we can generate LTE signals at any frequency band under 6GHz without the limitation of using certain discrete BPF like the RF-DAC platform. We performed a set of DPD tests at different bands. Particularly, we performed a test on a 10W PA (CGH40010f) at 4GHz using a 20MHz LTE signal. Without any analog mixing stage, our proposed undersampling DPD approach achieved satisfactory -54 dBc ACPR performance using a single-ADC based feedback path running at 61.44MSPS, as shown in Fig. 6.18. In addition, we provided AM-AM, AM-PM curves, as shown in Fig. 6.19, for illustration of the linearization performance as well.

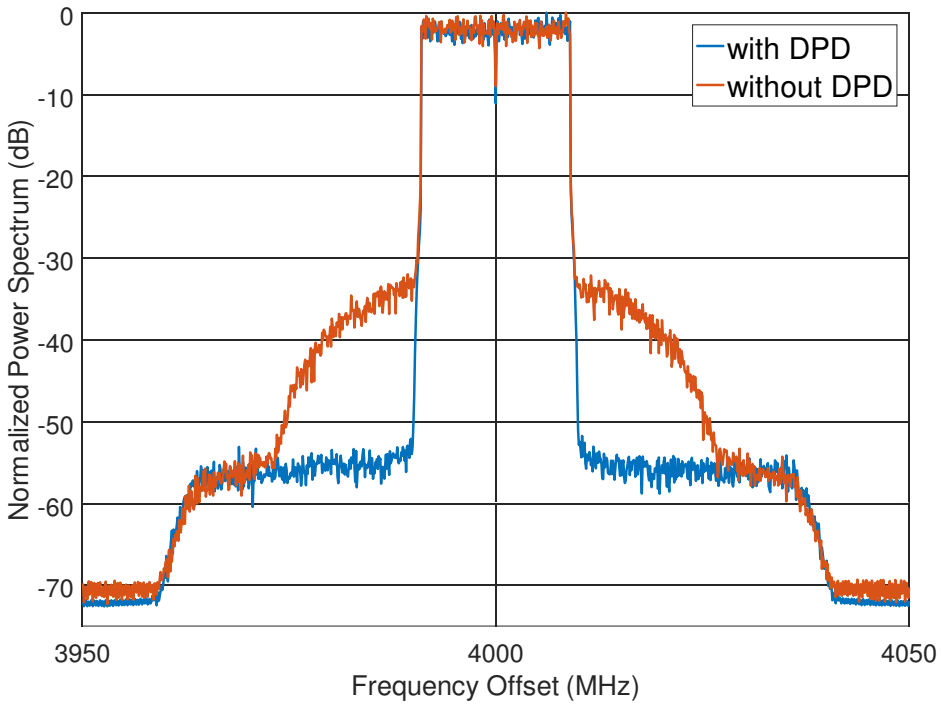


Figure 6.18: Spectra of PA outputs before and after proposed DPD using a single-carrier LTE (20MHz) at 4GHz.

6.5 Conclusion

This chapter presented a novel compact undersampling digital predistortion-based PA linearization solution (including both hardware architecture and DPD algorithm) for single-chain wideband and multi-band RF transmitter. We propose a single-ADC based DPD observation receiver architecture, i.e., using only one ADC with certain analog filters to replace conventional DPD feedback paths that usually include mixers or analog demodulators, LO, and middle-to-high range ADCs. By addressing several parameters of ADC and corresponding DPD algorithm, we achieved multi-GHz DPD bandwidth coverage, and demonstrated the satisfactory linearization performance of multi-band and wideband RF transmitter applications via extensive experimental tests. Particularly, we evaluated three application scenarios: a) 3-band carrier aggregated LTE sig-

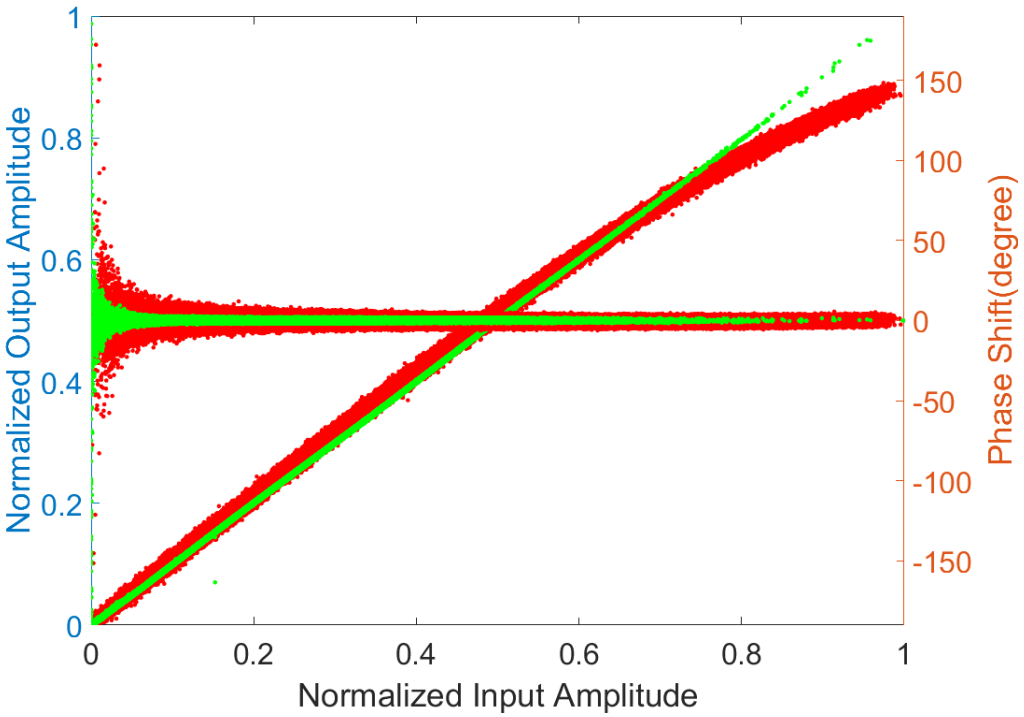


Figure 6.19: AM-AM and AM-PM of PA output of 20 MHz bandwidth signal with and without proposed DPD solution at 4GHz.

nals (each one occupying 20MHz instantaneous bandwidth), the linearization performance achieved 50 dBc ACPR with more than 1GHz bandwidth using ADC running at 76.8 MSPS and 61.44 MSPS; b) single-band LTE signal with continuous 100 MHz instantaneous bandwidth, the linearization performance achieved -48 dBc ACPR using 61.44 MSPS ADC, and c) single-band LTE signal with continuous 20MHz bandwidth at 4GHz, the linearization performance achieved -54 dBc ACPR using 61.44 MSPS ADC. Furthermore, the proposed undersampling ADC based DPD feedback path can be flexibly tailored by employing different speed track-and-hold circuits for different applications, meeting various requirements of system performance and cost.

Chapter 7

Conclusion

7.1 Summary of Thesis Achievements

The wideband capturing and large-matrix multiplications lead to DPD added power consumption. Direct use of lower sampling-rate and less number of feedback samples will increase generalization errors in the DPD training procedure. To answer these challenges, in this thesis we firstly proposes a theory - information carried by modulated signals, to connect the statistic properties of modulated signals with the generalization ability of memory polynomial based models using LS-based algorithms. According to the proposed theory, the information carried by training samples, used for memory polynomial based model, is to a joint probability density function, which is foreknown with respect to PAPR and bandwidth of the transmitted modulated signal. Thus, a compressed sample selection method is proposed to reduce the number of training samples by only selecting the minimal samples which satisfy the foreknown joint PDF. Furthermore, we prove that theoretically a memory polynomial based DPD can be performed using low sampling-rate of feedback signals. Consequently, the gener-

alized undersampling DPD is introduced in order to provide that the principle, implementation and practical concerns of DPD systems using undersampling ADCs. Finally, the multi-rate track-and-hold circuit with assisted algorithms is proposed to extend the observation bandwidth of DPD systems to multi-GHz.

This thesis aims to research the theory and techniques of DPD systems for reducing the DPD added algorithm power consumption and added circuitry power consumption. In this work, the number of training samples and complex multiplication operations required for coefficients estimation can be reduced by more than 10 times without additional calculation resource. Signal shaping techniques and measurement noise limit the minimum complexity reduction using the proposed method. On the other hand, the DPD observation path achieves 4 GHz bandwidth using an ADC running at 61.44 MSPS. The DPD added circuitry power consumption and hardware cost of observation path are relieved. To the best of our knowledge in the literature, it is the widest DPD observation path achieved so far using a real undersampling ADC. It could be attractive to the present and future wideband and low-power DPD systems, especially it is promising for the millimeter wave DPD application. In summary, in this thesis, we have studied the information carried by modulated training samples used for the least-square based DPD estimation. We have experimentally shown that the number and the sampling-rate of feedback samples can be dramatically reduced without loss of information for training. Moreover, the new concepts proposed in this thesis not only reduce the DPD added complexities, but also connect the statistical properties of modulated signals to the DPD training. These contributions are expected to make a broad impact on the DPD research and realization aspects.

7.2 Future Work

Running a feedback ADC at a slower throughput can be very advantageous. Since the acquisition period of the ADC is the most common area of the conversion cycle to read conversion data, a longer acquisition period will relax the requirements of the digital host including the interface with FPGA and the digital signal processing. Perhaps all of ADCs used in the observation path of modern DPD systems for the past ten years are pipeline ADCs. The pipeline ADC appears popular because of the high sampling-rate with high resolution. However, the power consumption of a pipeline ADC is considerably high even though running at a lower sampling-rate. Thus, employing the successive-approximation-register (SAR) ADC to the observation path of wideband DPD systems is very attractive, especially for the small cell and Massive MIMO system where the DPD added power consumption is more sensitive to the system efficiency. The SAR ADC generally requires significantly less silicon area and power consumption than an equivalent pipeline ADC. To date, the maximum sampling-rate of high-resolution SAR ADC is up to 5Msps, a three-stage multi-rate sample-and-hold (or track-and-hold) circuit is required to stepwise extend the capturing bandwidth to GHz. As shown in Fig. 7.1, a framework of an RF-sampling transceiver using RF DAC in the forward path and SAR ADC in the observation path is proposed as a future project.

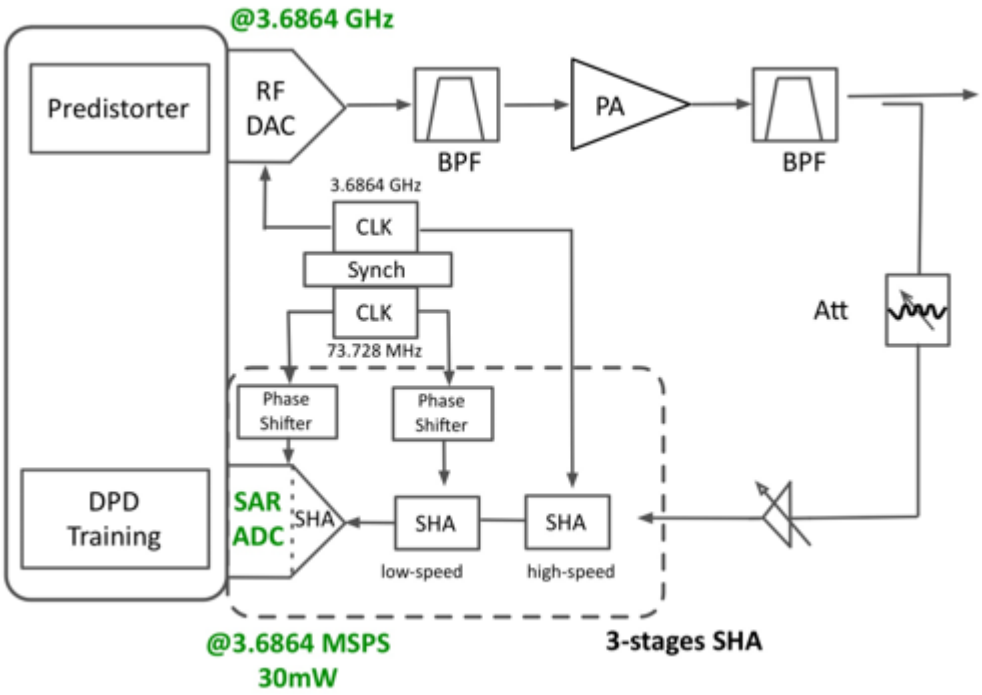


Figure 7.1: RF-sampling Transceiver with SAR ADC

Bibliography

[1] K. Chan, "Gc5325 envelope tracking," *Texas Instruments, Application Report SLWA058*, 2010.

[2] R. Darraji and F. M. Ghannouchi, "Digital doherty amplifier with enhanced efficiency and extended range," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 11, pp. 2898–2909, 2011.

[3] C. Yu, J. Xia, X.-W. Zhu, and A. Zhu, "Single-model single-feedback digital predistortion for concurrent multi-band wireless transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 7, pp. 2211–2224, 2015.

[4] Y. Liu, J. J. Yan, and P. M. Asbeck, "Concurrent dual-band digital predistortion with a single feedback loop," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, pp. 1556–1568, May 2015.

[5] T. W. Telecommunication/ICT, "Ict facts and figures 2017," *International Telecommunication Union*, 2017.

[6] J. Wood, *Behavioral modeling and linearization of RF power amplifiers*. Artech House, 2014.

[7] L. Guan and A. Zhu, "Green communications: Digital predistortion for wideband rf power amplifiers," *IEEE Microwave Magazine*, vol. 15, no. 7, pp. 84–99, 2014.

[8] C.-P. Liang, J.-h. Jong, W. E. Stark, and J. R. East, "Nonlinear amplifier effects in communications systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1461–1466, 1999.

[9] T. Edler, "Green base stationshow to minimize co2 emission in operator networks," in *Ericsson seminar, Bath Base Station Conf*, 2008.

[10] R. Pengelly, C. Fager, and M. Ozen, "Doherty's legacy: a history of the doherty power amplifier from 1936 to the present day," *IEEE Microwave Magazine*, vol. 17, no. 2, pp. 41–58, 2016.

[11] N. O. Sokal, "Rf power amplifiers, classes a through s-how they operate, and when to use each," in *Electronics Industries Forum of New England, 1997. Professional Program Proceedings*, pp. 179–252, IEEE, 1997.

[12] G. Wimpenny, "Improving multi-carrier pa efficiency using envelope tracking," *RF DesignLine*, 2008.

[13] G. Montoro, P. L. Gilabert, P. Vizarreta, and E. Bertran, "Slew-rate limited envelopes for driving envelope tracking amplifiers," in *2011 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications*, pp. 17–20, IEEE, 2011.

[14] W. H. Doherty, "A new high efficiency power amplifier for modulated waves," *Proceedings of the Institute of Radio Engineers*, vol. 24, no. 9, pp. 1163–1182, 1936.

[15] J. P. Martins, P. M. Cabral, N. B. Carvalho, and J. C. Pedro, "A metric for the quantification of memory effects in power amplifiers," *Microwave*

Theory and Techniques, IEEE Transactions on, vol. 54, no. 12, pp. 4432–4439, 2006.

[16] A. Zhu and T. J. Brazil, "Behavioral modeling of rf power amplifiers based on pruned volterra series," *IEEE Microwave and Wireless components letters*, vol. 14, no. 12, pp. 563–565, 2004.

[17] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microwave Magazine*, vol. 10, no. 7, pp. 52–64, 2009.

[18] M. Schetzen, "The volterra and wiener theories of nonlinear systems," 1980.

[19] M. Schetzen, "Nonlinear system modeling based on the wiener theory," *Proceedings of the IEEE*, vol. 69, no. 12, pp. 1557–1573, 1981.

[20] L. Ding, R. Raich, and G. T. Zhou, "A hammerstein predistortion linearization design based on the indirect learning architecture," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 3, pp. III–2689, IEEE, 2002.

[21] 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," *Signal Processing, IEEE Transactions on*, vol. 54, no. 10, pp. 3852–3860, 2006.

[22] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 58, no. 6, pp. 1510–1520, 2010.

[23] L. Ding, *Digital predistortion of power amplifiers for wireless applications*. PhD thesis, Georgia Institute of Technology, 2004.

[24] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *IEEE transactions on signal processing*, vol. 55, no. 1, pp. 120–133, 2007.

[25] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," pp. 309–313, 2008.

[26] Z. Yu and E. Zhu, "A comparative study of learning architecture for digital predistortion," in *Microwave Conference (APMC), 2015 Asia-Pacific*, vol. 1, pp. 1–3, IEEE, 2015.

[27] Y.-M. Zhu, "Generalized sampling theorem," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 39, no. 8, pp. 587–588, 1992.

[28] J. Tsimbinos and K. V. Lever, "Input nyquist sampling suffices to identify and compensate nonlinear systems," *Signal Processing, IEEE Transactions on*, vol. 46, no. 10, pp. 2833–2837, 1998.

[29] H. Koeppl and P. Singerl, "An efficient scheme for nonlinear modeling and predistortion in mixed-signal systems," *Circuits and Systems II: Express Briefs, IEEE Transactions on*, vol. 53, no. 12, pp. 1368–1372, 2006.

[30] P. Singerl and H. Koeppl, "A low-rate identification method for digital predistorters based on volterra kernel interpolation," *Analog Integrated Circuits and Signal Processing*, vol. 56, no. 1-2, pp. 107–115, 2008.

[31] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Open-loop digital predistorter for rf power amplifiers using dynamic deviation reduction-based volterra series," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 56, no. 7, pp. 1524–1534, 2008.

[32] Y. Liu, W. Pan, S. Shao, and Y. Tang, "A general digital predistortion architecture using constrained feedback bandwidth for wideband power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 63, pp. 1544–1555, May 2015.

[33] 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," 2014.

[34] Y. Ma, Y. Yamao, Y. Akaiwa, and K. Ishibashi, "Wideband digital predistortion using spectral extrapolation of band-limited feedback signal," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 61, no. 7, pp. 2088–2097, 2014.

[35] S. Bensmida, O. Hammi, A. Kwan, M. S. Sharawi, K. A. Morris, and F. M. Ghannouchi, "Extending the characterization bandwidth of dynamic nonlinear transmitters with application to digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 8, pp. 2640–2651, 2016.

[36] 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 Transactions on Microwave Theory and Techniques*, vol. 64, pp. 3528–3539, Nov 2016.

[37] Z. Wang, J. Dooley, K. Finnerty, and R. Farrell, "A low-rate identification scheme of high power amplifiers," in *Proceedings of the 17th Research Colloquium on Communications and Radio Science into the 21st Century*, Royal Irish Academy, 2014.

[38] Z. Wang, S. Ibrahim, H. Su, and R. Farrell, "Generalised digital predistortion of rf power amplifiers with low-rate feedback signal," in *2016 46th European Microwave Conference (EuMC)*, pp. 831–834, Oct 2016.

[39] H. Huang, P. Mitran, and S. Boumaiza, "Digital predistortion function synthesis using undersampled feedback signal," *IEEE Microwave and Wireless Components Letters*, vol. 26, pp. 855–857, Oct 2016.

[40] Z. Wang, L. Guan, and R. Farrell, "Compact undersampled digital predistortion for flexible single-chain multi-band rf transmitter," in *IEEE MTT-S International Microwave Symposium 2017 (IMS2017)*, 2017.

[41] Z. Wang, L. Guan, and R. Farrell, "Undersampling observation-based compact digital predistortion for single-chain multiband and wideband direct-to-rf transmitter," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 12, pp. 5274–5283, 2017.

[42] W. A. Frank, "Sampling requirements for volterra system identification," *IEEE Signal Processing Letters*, vol. 3, no. 9, pp. 266–268, 1996.

[43] R. N. Braithwaite, "Wide bandwidth adaptive digital predistortion of power amplifiers using reduced order memory correction," in *Microwave Symposium Digest, 2008 IEEE MTT-S International*, pp. 1517–1520, IEEE, 2008.

[44] O. Hammi, A. Kwan, S. Bensmida, K. A. Morris, and F. M. Ghannouchi, "A digital predistortion system with extended correction bandwidth with application to lte-a nonlinear power amplifiers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 12, pp. 3487–3495, 2014.

[45] T. Wang, P. L. Gilabert, and G. Montoro, "Under-sampling effects and computational cost reduction in rf power amplifier behavioral modeling,"

in *Microwave Integrated Circuits Conference (EuMIC), 2015 10th European*, pp. 57–60, IEEE, 2015.

[46] O. Andersen, N. Bjorsell, and N. Keskitalo, "A test-bed designed to utilize zhu's general sampling theorem to characterize power amplifiers," in *Instrumentation and Measurement Technology Conference, 2009. I2MTC'09. IEEE*, pp. 201–204, IEEE, 2009.

[47] D. H. Wisell, "Exploring the sampling rate requirements for behavioural amplifier modelling," in *18th IMEKO World Congress 2006: Metrology for a Sustainable Development*, 2006.

[48] N. Guan, N. Wu, and H. Wang, "Digital predistortion of wideband power amplifier with single undersampling adc," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 11, pp. 1016–1018, 2017.

[49] S. A. Bassam, M. Helaoui, and F. M. Ghannouchi, "2-d digital predistortion (2-d-dpd) architecture for concurrent dual-band transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, pp. 2547–2553, Oct 2011.

[50] S. A. Bassam, A. Kwan, W. Chen, M. Helaoui, and F. M. Ghannouchi, "Subsampling feedback loop applicable to concurrent dual-band linearization architecture," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, pp. 1990–1999, June 2012.

[51] Y. J. Liu, W. Chen, J. Zhou, B. H. Zhou, and F. M. Ghannouchi, "Digital predistortion for concurrent dual-band transmitters using 2-d modified memory polynomials," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 281–290, Jan 2013.

[52] M. Younes, A. Kwan, M. Rawat, and F. M. Ghannouchi, "Linearization of concurrent tri-band transmitters using 3-d phase-aligned pruned volterra model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 4569–4578, Dec 2013.

[53] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-band complexity-reduced generalized-memory-polynomial power-amplifier digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, pp. 1763–1774, June 2016.

[54] P. L. Gilabert and G. Montoro, "3-d distributed memory polynomial behavioral model for concurrent dual-band envelope tracking power amplifier linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 638–648, 2015.

[55] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 5, pp. 1880–1890, 2017.

[56] R. Cudeck, "Analysis of correlation matrices using covariance structure models.," *Psychological Bulletin*, vol. 105, no. 2, p. 317, 1989.

[57] L. Guan and A. Zhu, "Optimized low-complexity implementation of least squares based model extraction for digital predistortion of rf power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 60, no. 3, pp. 594–603, 2012.

[58] D. A. Schum, *The evidential foundations of probabilistic reasoning*. Northwestern University Press, 1994.

[59] J. Goerlich, D. Bruckner, A. Richter, O. Strama, R. Thoma, and U. Trautwein, "Signal analysis using spectral correlation measure-

ment," in *Instrumentation and Measurement Technology Conference, 1998. IMTC/98. Conference Proceedings. IEEE*, vol. 2, pp. 1313–1318, IEEE, 1998.

[60] I. Miller, "Probability, random variables, and stochastic processes," 1966.

[61] R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 5, pp. 1468–1479, 2004.

[62] S. Wu, J. Zhai, Z. Yu, and N. Zhang, "A compressed sensing method for pruning concurrent dual-band power amplifier models," in *2018 IEEE MTT-S International Wireless Symposium (IWS)*, pp. 1–3, IEEE, 2018.

[63] P. L. Gilabert, G. Montoro, T. Wang, M. N. Ruiz, and J. A. García, "Comparison of model order reduction techniques for digital predistortion of power amplifiers," in *Microwave Conference (EuMC), 2016 46th European*, pp. 182–185, IEEE, 2016.

[64] M. V. Amiri, S. Bassam, M. Helaoui, and F. Ghannouchi, "New order selection technique using information criteria applied to siso and mimo systems predistortion," *International Journal of Microwave and Wireless Technologies*, vol. 5, no. 2, pp. 123–131, 2013.

[65] W. Chen, S. Zhang, Y.-J. Liu, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Efficient pruning technique of memory polynomial models suitable for pa behavioral modeling and digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 10, pp. 2290–2299, 2014.

[66] P. L. Gilabert, G. Montoro, D. Lopez, N. Bartzoudis, E. Bertran, M. Payaro, and A. Hourtane, "Order reduction of wideband digital predistorters us-

ing principal component analysis," in *Microwave Symposium Digest (IMS), 2013 IEEE MTT-S International*, pp. 1–7, IEEE, 2013.

[67] J. Wood, "Digital pre-distortion of rf power amplifiers: progress to date and future challenges," in *Microwave Symposium (IMS), 2015 IEEE MTT-S International*, pp. 1–3, IEEE, 2015.

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

[69] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low computational complexity digital predistortion based on direct learning with covariance matrix," *IEEE Transactions on Microwave Theory and Techniques*, 2017.

[70] L. D. Brown and J. G. Hwang, "How to approximate a histogram by a normal density," *The American Statistician*, vol. 47, no. 4, pp. 251–255, 1993.

[71] X. Yu and H. Jiang, "Digital predistortion using adaptive basis functions," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, pp. 3317–3327, Dec 2013.

[72] J. Reina-Tosina, M. Allegue-Martínez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 745–753, 2015.

[73] D. Lpez-Bueno, Q. A. Pham, G. Montoro, and P. L. Gilabert, "Independent digital predistortion parameters estimation using adaptive principal compo-

nent analysis," *IEEE Transactions on Microwave Theory and Techniques*, pp. 1–9, 2018.

[74] S. A. Bassam, W. Chen, M. Helaoui, F. M. Ghannouchi, and Z. Feng, "Linearization of concurrent dual-band power amplifier based on 2d-dpd technique," *IEEE Microwave and Wireless Components Letters*, vol. 21, no. 12, pp. 685–687, 2011.

[75] Y.-J. Liu, W. Chen, J. Zhou, B.-H. Zhou, and F. M. Ghannouchi, "Digital predistortion for concurrent dual-band transmitters using 2-d modified memory polynomials," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 281–290, 2013.

[76] A. Prata, D. C. Ribeiro, P. M. Cruz, A. S. Oliveira, and N. B. Carvalho, "Rf subsampling feedback loop technique for concurrent dual-band pa linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 12, pp. 4174–4182, 2016.

[77] Y. Liu, J. J. Yan, and P. M. Asbeck, "Concurrent dual-band digital predistortion with a single feedback loop," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 5, pp. 1556–1568, 2015.

[78] H. Qian, S. Yao, H. Huang, X. Yang, and W. Feng, "Low complexity coefficient estimation for concurrent dual-band digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 10, pp. 3153–3163, 2015.

[79] Z. Wang, J. Dooley, K. Finnerty, and R. Farrell, "Selection of compressed training data for rf power amplifier behavioral modeling," in *Proceedings of the 10th European Microwave Integrated Circuits Conference*, European Microwave Association, 2015.

[80] P. D. Sutton, B. Özgül, and L. Doyle, "Cyclostationary signatures for lte advanced and beyond," *Physical Communication*, vol. 10, pp. 179–189, 2014.

[81] M. A. Hussein, V. A. Bohara, and O. Venard, "On the system level convergence of ila and dla for digital predistortion," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, pp. 870–874, IEEE, 2012.

[82] J. Chani-Cahuana, M. Ozen, C. Fager, and T. Eriksson, "Digital predistortion parameter identification for rf power amplifiers using real-valued output data," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2017.

[83] P. T. Congdon, P. Mohapatra, M. Farrens, and V. Akella, "Simultaneously reducing latency and power consumption in openflow switches," *IEEE/ACM Transactions on Networking (TON)*, vol. 22, no. 3, pp. 1007–1020, 2014.

[84] L. Guan and A. Zhu, "Simplified dynamic deviation reduction-based volterra model for doherty power amplifiers," in *Integrated Nonlinear Microwave and Millimetre-Wave Circuits (INMMIC), 2011 Workshop on*, pp. 1–4, IEEE, 2011.

[85] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited volterra series-based digital predistortion for wideband rf power amplifiers," 2012.

[86] L. Guan, *FPGA-based Digital Convolution for Wireless Applications*. Springer, 2017.

[87] A. Devices, "Ultra-wideband 4 gs/s track-and-hold amplifier, dc - 18 ghz," *HMC661LC4B datasheet*.