

Novel Predistortion System for 4G/5G Small-Cell and Wideband Transmitters

by

Hai Huang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2020

© Hai Huang 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Thomas Eriksson
Professor, Dept. of Signal Processing,
Chalmers University of Technology

Supervisors: Slim Boumaiza
Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Peter Levine
Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Internal Members: David Narin
Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Patrick Mitran
Professor, Dept. of Electrical and Computer Engineering,
University of Waterloo

Internal-External Member: Baris Fidan
Professor, Dept. of Mechanical and Mechatronics Engineering,
University of Waterloo

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

To meet the growing demand for mobile data, various technologies are being introduced to wireless networks to increase system capacity. On one hand, large number of small-cell base stations are adopted to serve the reduced cell size; on the other hand, millimeter wave (mm-wave) systems with large antenna arrays that transmit ultra-wideband signals are expected in fifth generation (5G) networks. Power amplifiers (PAs), responsible for boosting the radio frequency (RF) signal power, are the most critical components in base station transmitters, and dominate the overall efficiency and linearity of the system. The design challenges to balance the contradictory requirements of efficiency and linearity of the PAs are usually addressed by linearization techniques, particularly the digital predistortion (DPD) system. However, existing DPD solutions face increasing difficulties keeping up with new developments in base station technologies.

When considering sub-6 GHz small-cell base station transmitters, analog and RF predistortion techniques have recently received renewed attention due to their inherent low power nature. Their achievable linearization capacity is significantly limited, however, largely by their implementation complexity in realizing the needed predistortion models in analog circuitry. On the other hand, despite significant developments in DPD models for wideband signals, the implementations of such DPD models in practical hardware have received relatively little attention. Yet the conventional implementation of a DPD engine is limited by the maximum clock frequency of the digital circuitry employed and cannot be scaled to satisfy the growing bandwidth of transmitted signals for 5G networks. Furthermore, both analog and digital solutions require a transmitter-observation-receiver (TOR) to capture the PA outputs, necessitates the use of analog-to-digital converters (ADCs) whose complexity and power consumption increase with signal bandwidth. Such trend is not scalable for future base stations, and new innovations in feedback and training methods are required. This thesis presents a number of contributions to address the above identified challenges.

To reduce the power overhead of the linearization system, a digitally-assisted analog-RF predistortion (DA-ARFPD) system that uses a novel predistortion model is introduced. The proposed finite-impulse-response assisted envelope memory polynomial (FIR-EMP) model allows for a reduction of hardware implementation complexity while maintaining good linearization capacity and low power overhead. A two-step small-signal-assisted parameter identification (SSAPI) algorithm is devised to estimate the parameters of the two main blocks of the FIR-EMP model, such that the training can be completed efficiently. A DA-ARFPD test bench has been built, which incorporates major RF components, to assess the validity of the proposed FIR-EMP scheme and the SSAPI algorithm. Measurement

results show that the proposed FIR-EMP model with SSAPI algorithm can successfully linearize multiple PAs driven with various wideband and carrier-aggregated signals of up to 80 MHz modulation bandwidths for sub-6 GHz systems.

Next, a hardware-efficient real-time DPD system with scalable linearization bandwidth for ultra-wideband 5G mm-wave transmitters is proposed. It uses a novel parallel-processing DPD engine architecture to process multiple samples per clock cycle, overcomes the linearization bandwidth limit imposed by the maximum clock rate of digital circuits used in conventional DPD implementation. Potentially unlimited linearization bandwidth could be achieved by using the proposed system with current digital circuit technologies. The linearization performance and bandwidth scalability of the proposed system is demonstrated experimentally using a silicon-based Doherty PA (DPA) with 400 MHz wideband signal operating at 28 GHz, and over-the-air measurements using a 64-element beamforming array with 800 MHz wideband signal, also at 28 GHz. The proposed DPD system achieves over 2.4 GHz linearization bandwidth using only a 300 MHz core clock for the digital circuits.

Finally, to reduce the power consumption and cost of the TOR, a new approach to train the predistorter using under-sampled feedback signal is presented. Using aliased samples of the PA's output captured at either baseband or intermediate frequency (IF), the proposed algorithm is able to compute the coefficients of the predistortion engine to linearize the PA using a direct learning architecture. Experimentally, both the baseband and IF schemes achieve linearization performance comparable to a full-rate system. Implemented together with a parallel-processing based DPD engine on a field-programmable gate array (FPGA) based system-on-chip (SOC), the proposed feedback and training solution achieves over 2.4 GHz linearization bandwidth using an ADC operating at a clock rate of 200 MHz. Its performance is demonstrated experimentally by linearizing a silicon DPA with 200 MHz and 400 MHz signals in conductive measurements, and a 64-element beamforming array with 400 MHz and 800 MHz signals in over-the-air testing.

Acknowledgements

I would like to express my sincere thanks to all of those who helped me to get through this challenging PhD journey. I would like to give my greatest gratitude to professor Slim Boumaiza who has supported me and guided me to go through the ups and downs along the way. He gave me the opportunities to engage in these wonderful projects and explore different aspects of the research. I would also like to thank professor Peter Levine for his support and guidance. In addition, I would like to thank my committee members for reviewing my thesis and for their valuable feedback.

Throughout the past years, I am grateful to have had wonderful friends in the EmRG group. It has been a great experience working with them and learning from them. They also gave me great encouragement and made my life colorful and enjoyable. I would also like to take this opportunity to thank all my friends in Canada for their support and companionship.

Finally, I would like to thank my parents for their continuous support; without them I could not have reached where I am today.

Table of Contents

List of Figures	x
List of Tables	xiv
List of Acronyms	xv
1 Introduction	1
1.1 Motivation	1
1.2 Thesis Objective	4
1.3 Thesis Outline	5
2 Background and Literature on Power Amplifier Linearization	7
2.1 Introduction	7
2.2 Power Amplifier Efficiency	8
2.2.1 Conventional Power Amplifier Efficiency	10
2.2.2 Backoff Efficiency Enhancement Techniques	11
2.2.3 Other Efficiency Enhancement Techniques	13
2.3 Power Amplifier Linearity	16
2.3.1 Sources of Power Amplifier Distortion	18
2.3.2 Modelling of Power Amplifier Nonlinearity	20
2.3.3 Linearity versus Efficiency Trade-off	22

2.4	Linearization Techniques	23
2.4.1	Feedback Techniques	24
2.4.2	Feedforward Techniques	25
2.4.3	Predistortion Techniques	26
2.5	Literature Review on Predistortion Techniques	27
2.5.1	Overview of Digital Predistortion Techniques	28
2.5.2	Architectures of Predistortion Engine	30
2.5.3	Predistortion Function Formulation	34
2.5.4	Feedback and Training	37
2.5.5	Hardware Implementation of Predistortion System	38
2.6	Summary	39
3	Digitally Assisted Analog Radio Frequency Predistortion System	41
3.1	Introduction	41
3.2	Linear Filter Assisted Envelope Memory Polynomial Model Overview	42
3.3	FIR-EMP Parameter Identification Algorithm	46
3.4	Validation and Measurement Results	52
3.5	Conclusion	63
4	Real-time Digital Predistortion Hardware Architecture for Enabling Wide-band Linearization of 5G Transmitters	64
4.1	Introduction	64
4.2	Parallel-Processing-Based Digital Predistortion Engine	65
4.3	Measurement Results	73
4.4	Conclusion	81

5	Predistortion Function Synthesis using Under-sampled Feedback Signals	82
5.1	Introduction	82
5.2	Direct Learning Framework	83
5.3	DPD Function Synthesis Using Under-sampled Feedback Signal	85
5.4	DPD Function Synthesis Using Under-sampled Feedback Signal at IF	87
5.4.1	Hardware Optimization of Under-sampled Feedback	90
5.5	Delay Alignment	91
5.6	Validation and Measurement Results	93
5.6.1	Under-sampling using Baseband Scheme	93
5.6.2	Under-sampling using IF Scheme	95
5.7	Conclusion	101
6	Conclusion	102
6.1	Summary of Contributions and Publications	104
6.2	Future Work	105
	References	107

List of Figures

2.1	Block diagram of a conventional PA.	8
2.2	DE of ideal class A and class B PAs with the PDF of a long term evolution advanced (LTE-A) signal with 7.6 dB PAPR.	10
2.3	Block diagram of an ET PA.	12
2.4	Block diagram of a DPA.	12
2.5	Block diagram of an outphasing PA.	14
2.6	Block diagram of digital PA (a) and power mixer (b).	15
2.7	Constellation of 16-QAM (a) and error vector (b).	17
2.8	Illustration of PA spectra regrowth.	18
2.9	Gain (a) and phase (b) distortion of PA under modulated signal, with static distortion highlighted.	21
2.10	Linearity versus efficiency of PA.	23
2.11	Block diagram of a feedback linearization technique	24
2.12	Block diagram of a feedforward linearization technique.	25
2.13	The principle of predistortion linearization techniques.	26
2.14	Block diagram of a predistortion system	28
2.15	Classification of predistortion engines.	31
2.16	Block diagram of a digital/RF predistortion system.	32
2.17	Block diagram of an analog/RF predistortion system.	33
2.18	Block diagram of the indirect learning (a) and direct learning (b) architecture.	37

3.1	Block diagram of the analog/RF predistortion system.	42
3.2	Block diagram of the memory polynomial (MP) function when applied to the ARFPD.	44
3.3	AM/AM and AM/PM characteristics, measured and modelled using the envelope memory polynomial (EMP) model of a gallium nitride (GaN) Doherty PA (DPA) driven with a 20 MHz WCDMA signal at 2 GHz.	45
3.4	Block diagram of the proposed FIR filter assisted EMP function when applied to the DA-ARFPD.	47
3.5	Training scheme for identifying the coefficients of the FIR-EMP predistorter.	48
3.6	Steps of the proposed SSAPI algorithm for the FIR-EMP predistorter (a) FIR parameter identification using the forward model of the PA (b) EMP parameter identification using the intermediate output of the FIR block.	50
3.7	Modelling accuracy vs. the number of iterations for the quasi-Newton non-linear optimization and the proposed SSAPI algorithm.	52
3.8	Block diagram of the DA-ARFPD engine test bench and the PAs under test.	53
3.9	Photograph of the DA-ARFPD measurement setup.	53
3.10	(a) AM/AM and (b) AM/PM characteristics for PA1 without predistortion (red) and with the proposed FIR-EMP in digitally-assisted ARFPD (DA-ARFPD) test bench (blue) of a 30 MHz wideband code division multiple access (WCDMA) signal.	58
3.11	(a) AM/AM and (b) AM/PM characteristics for PA2 without predistortion (red) and with the proposed FIR-EMP in DA-ARFPD test bench (blue) of a 30 MHz WCDMA signal.	59
3.12	Output spectra of PA1 driven with 30 MHz bandwidth signal.	60
3.13	Output spectra of PA2 driven with 30 MHz bandwidth signal.	61
3.14	Output spectra of PA2 driven by 80 MHz signal.	62
4.1	Block diagram of a typical DPD system.	65
4.2	Data flow diagram of conventional serial processing DPD engine, assuming memory depth $M = 5$	67
4.3	Block diagram of proposed system with parallel processing DPD engine, having parallel acceleration factor F and under-sampling feedback signal sampled at f_{RX} and sampled $\tilde{y}'_m = \tilde{y}(m \times R/f_s)$	68

4.4	Data flow diagram of proposed parallel processing DPD engine, assuming memory depth $M = 5$ and parallel acceleration factor $F = 4$	69
4.5	Detailed implementation of a section of the parallel-processing DPD engine illustrating the pipelined polynomial generation, the cross-bar, and one branch of the coefficient application stage.	70
4.6	Block diagram of a sub-section of a simple cross-bar block with $F = 2$ and $M = 5$, with routing for each branch shown in different color.	71
4.7	High level block diagram of DSP48 unit in Xilinx field-programmable gate array (FPGA).	73
4.8	Measurement setup of proposed DPD system – linearizing a silicon DPA in a probe station.	75
4.9	Output spectra of mm-wave silicon DPA in conducted measurement setup operating at 28 GHz using signals with 200 MHz bandwidths.	76
4.10	Output spectra of mm-wave silicon DPA in conducted measurement setup operating at 28 GHz using signals with 400 MHz bandwidths.	77
4.11	Measurement setup of proposed DPD system – linearizing a 64-element mm-wave beamforming array with over-the-air (OTA) receiving antenna.	79
4.12	Output spectra of mm-wave 64-element beamforming array in the OTA measurement setup operating at 28 GHz using signals with 400 MHz bandwidth.	80
4.13	Output spectra of mm-wave 64-element beamforming array in the OTA measurement setup operating at 28 GHz using signals with 800 MHz bandwidth.	80
5.1	Block diagram of the proposed under-sampling feedback system with a feedback signal sampled at significantly reduced rate of f_{RX} . The conventional system corresponds to the special case where $f_{RX} = f_s$	86
5.2	Measurement setup of the proposed baseband under-sampling scheme.	94
5.3	Spectrum of the PA output driven with the LTE-A signal.	95
5.4	Spectrum of the PA output driven with the CA signal.	96
5.5	Measurement setup of proposed DPD system – linearizing a silicon DPA in a probe station.	97
5.6	Output spectra of mm-wave silicon DPA in conducted measurement setup operating at 28 GHz using signals with 400 MHz bandwidths trained with under-sampling TOR.	98

5.7	Measurement setup of proposed DPD system – linearizing a 64-element mm-wave beamforming array with OTA receiving antenna.	99
5.8	Output spectra of mm-wave 64-element beamforming array in the OTA measurement setup operating at 28 GHz using signals with 800 MHz bandwidth trained with under-sampling TOR.	100

List of Tables

1.1	Typical power consumption of DPD and base stations in 4G networks . . .	3
2.1	Summary of the different classes of predistortion engine	31
3.1	Summary of the Linearization Performance of PA1 Driven by Wideband Modulated Signals	55
3.2	Summary of the Linearization Performance of PA2 Driven by Wideband Modulated Signals	56
4.1	Resource Utilization of Proposed DPD Engine (with nonlinearity order $N = 9$, memory depth $M = 5$ and parallel acceleration factor $F = 8$.)	74
4.2	Summary of Linearization Performance of millimeter wave (mm-wave) Silicon DPA and mm-wave 64-Element Beamforming Array (operating at 28 GHz).	81
5.1	Measurement Results of Baseband Based Training with Under-sampled Feedback Signal	93
5.2	Measurement Results of IF Based Training with Under-sampled Feedback Signal	100

List of Acronyms

4G forth generation.

5G fifth generation.

ACLR adjacent channel leakage ratio.

ACPR adjacent channel power ratio.

ADC analog-to-digital converter.

ANN artificial neural network.

ARFPD analog/RF predistortion.

CA carrier aggregation.

CMOS complimentary metal-oxide-semiconductor.

CRV complexity reduced Volterra.

CW continuous wave.

DA-ARFPD digitally-assisted ARFPD.

DAC digital-to-analog converter.

DC direct current.

DE drain efficiency.

DPA Doherty PA.

DPD digital predistortion.

DSP digital signal processor.

DUT device under test.

ECRV envelope complexity reduced Volterra.

EMP envelope memory polynomial.

ET envelope tracking.

EVM error vector magnitude.

FPGA field-programmable gate array.

GaN gallium nitride.

GMP generalized memory polynomial.

IF intermediate frequency.

LINC linear amplification with nonlinear components.

LMS least mean square.

LS least square.

LTE-A long term evolution advanced.

LUT lookup table.

MIMO multiple-input multiple-output.

mm-wave millimeter wave.

MP memory polynomial.

NMSE normalized mean square error.

NR new radio.

OFDM orthogonal frequency division multiplexing.

OTA over-the-air.

PA power amplifier.

PAE power added efficiency.

PAPR peak-to-average power ratio.

QAM quadrature amplitude modulation.

RF radio frequency.

RLS recursive least square.

SISO single-input single-output.

SNIR signal to interference and noise.

SOC system-on-chip.

SOI silicon-on-insulator.

TOR transmitter-observation-receiver.

WCDMA wideband code division multiple access.

Chapter 1

Introduction

1.1 Motivation

The world has become a much more connected place with over 8 billion mobile subscriptions around the globe by the end of 2019 which is expected to continue increasing in the following years. The wireless networks have also changed from voice-centered service in the second generation era to data-centered mobile broadband that supports services such as mobile payment, high definition video streaming and many others that have fundamentally changed people's way of life. Mobile data traffic is expected to continue its exponential growth, forecast to reach 160 exabytes per month in 2025 [1]. Such demand for high speed mobile data is one of the key driving forces behind the rapid development of mobile technology to provide mobile broadband service to billions of users. Besides improving the capacity of current fourth generation (4G) long term evolution advanced (LTE-A) networks, the upcoming fifth generation (5G) networks that have already begun rapid deployment in many countries around the globe aim to drastically increase the speed of data transfer with data rates expected to reach 1 Gbps in the near future.

Various technologies have been introduced to increase system capacity, such as reducing cell size by using small-cell base stations, improving spectrum efficiency with high order modulation schemes, utilising spatial multiplexing by deploying multiple-input multiple-output (MIMO) systems, as well as increasing signal bandwidths through carrier aggregation (CA) and the use of millimeter wave (mm-wave) spectrum. However, such technologies impose demanding and contradictory requirements in terms of linearity and efficiency for radio frequency (RF) front-ends, particularly the RF power amplifier (PA) that consumes the largest amount of power and contributes most of the distortion in a transmitter. Lin-

earization techniques, particularly digital predistortion (DPD) systems, have become a necessity to improve the linearization and efficiency of RF transmitters.

One of the most effective ways to increase the system capacity of current networks is to reduce the cell size, so that each base station is serving a smaller number of users. This enables resources such as spectrum or time slots to be reused, effectively increasing the available capacity to each user, hence achieving high total capacity for the system. The distances between base stations have decreased from tens of kilometers for macro-cells to a few hundred meters or below for small-cells such as pico- or femto-cells, and the transmitted power of base stations has decreased from hundreds of Watts to a few Watts or less. With shrinking cells each covering a much smaller area, more base stations are needed to ensure coverage of the network. The number of small-cell base stations, which has already surpassed that of macro-cell base stations, is expected to increase rapidly [2]. The energy efficiency of the low-power small-cell base stations is critical to the overall power consumption of the wireless network.

At the same time, complex modulation schemes such as 64 quadrature amplitude modulation (QAM) and sophisticated access technologies such as orthogonal frequency division multiplexing (OFDM) are used to increase the spectrum efficiency of the channel. However, such schemes present difficult design challenges to the transmitter front-end due to the characteristics of the transmitted signal and the stringent requirements on signal quality. The high peak-to-average power ratio (PAPR) of the transmitted signals require innovative designs for efficient transmitters. PA topologies such as Doherty and envelope tracking (ET) are used to improve both the peak and backoff efficiency, at the cost of linearity due to their complex architecture. Meanwhile, the stringent requirements on signal quality, typically measured by error vector magnitude (EVM), adjacent channel leakage ratio (ACLR) or adjacent channel power ratio (ACPR), mean sophisticated linearization solutions are needed to achieve the required linearity. However, as the output power of each base station is reduced, the overhead of the linearization system becomes more significant. Table 1.1 summarizes the typical power consumption of different DPD components [3], as well as the power consumption and operation range of various base stations in 4G wireless networks. It is clear that conventional linearization solutions are not suitable for small-cell base stations, as the power overhead becomes a significant factor for the overall system, compromising their practicality.

On the other hand, the upcoming 5G networks will dramatically increase the available transmission bandwidth of the signal to each user to support much higher data rates. CA technology that groups multiple frequency blocks to the same user to provide wider transmission bandwidths has already been introduced in the current LTE-A networks, with up to 100 MHz total bandwidth allowed. With 5G new radio (NR), the transmission

Table 1.1: Typical power consumption of DPD and base stations in 4G networks

DPD Component	Power Consumption	Base Station Type	Typical Cell Radius	Typical PA Power Range
DPD Engine	1.5 W	Macro	> 1 km	20 W – 160 W
Clock Gen	0.4 W	Micro	250 m – 1 km	2 W – 20 W
Adaptation	0.6 W	Pico	100 m – 300 m	0.25 W – 2 W
Feedback	0.8 W	Femto	10 m – 50 m	0.01 W – 0.2 W

bandwidth is expected to growth significantly to reach several hundreds of megahertz, taking advantage of the unused spectrum in the mm-wave frequency. Massive MIMO systems made up of large antenna arrays that integrate a large number of antennas, PAs and other RF elements will be the key technology to support efficient transmission of such signals, as they allow beamforming that focuses the transmitted energy into a narrow beam to overcome the high path losses at such frequencies. However, the development of such massive MIMO systems is often hindered by poor efficiency and thermal issues. These issues are mainly attributable to the low efficiency of RF PAs in the transmission array as they must operate at significant backoff power levels to ensure linearity when transmitting wideband modulated signals. Yet, given the large number of PAs and close form factor, conventional DPD solutions designed for single PAs are usually not practical. Recent studies [4, 5] have focused on linearizing the entire array instead of individual PA elements. They have demonstrated that, in the case of RF/hybrid beamforming arrays, a single-input single-output (SISO) DPD model can be used to linearize the main-beam signal using a far-field probing receiving antenna, making DPD a viable solution for mm-wave systems.

Yet, two challenges exist to the practical and energy efficient realization of a DPD engine for wideband signals. First, the intermodulation caused by the nonlinear behaviour of the PA results in a bandwidth expansion of its output, and the predistorted signal required to compensate for such distortion also needs to occupy a wider bandwidth, typically 5x the signal bandwidth. As a result, the rate of the DPD engine needs to scale up proportionally. However, the highest achievable rate is limited by the maximum frequency attainable in the signal processing hardware. Second, the feedback path required to capture the PA distortion also experiences bandwidth expansion, necessitating a wideband feedback path and high-speed analog-to-digital converters (ADCs). To support 5G NR mm-wave communication signals with modulation bandwidths of several hundred of mega-

hertz, conventional DPD implementations would need digital circuits and data converters supporting bandwidths in the gigahertz. Such requirements render the hardware implementation of real-time MIMO DPD difficult, and ultimately can result in major energy and cost overhead in the underlying hardware.

1.2 Thesis Objective

The above discussion reveals major challenges in applying existing linearization solutions to support further improvements in system capacity for both current LTE-A and future 5G wireless communication systems. This thesis aims to address these problems through the following objectives:

1. Develop a low power analog/RF predistortion (ARFPD) architecture suitable for hardware implementation, and its corresponding coefficient update algorithm, for sub-6 GHz small-cell base stations. The proposed solution will take advantage of the low power feature of analog/RF circuits, with novel predistortion engine models that reduce the hardware complexity of the system while maintaining the same linearization capacity signals with moderate bandwidth.
2. Develop a real-time DPD hardware architecture for the linearization of 5G transmitters with ultra-wideband modulation signals found in 5G sub-6 GHz and mm-wave applications. The proposed solution will use a novel parallel-processing DPD engine architecture to overcome the maximum linearization bandwidth limit imposed by the maximum clock frequency of the digital circuits and support unlimited signal bandwidth.
3. Develop a low power and low cost transmitter-observation-receiver (TOR) and its accompanying algorithm to identify the coefficients of the predistortion system for adaptive control that is common to both analog and digital predistortion systems. The proposed technique will use an under-sampling scheme and direct learning algorithm that allows the adaptive update of the DPD coefficients independently of the DPD engine linearization bandwidth with a low sampling ADC in order to reduce the power consumption and complexity of the TOR.

1.3 Thesis Outline

Chapter 2 discusses the efficiency-linearity trade-off of PAs and provides a survey of PA linearization techniques and recent research work. It shows that linearization systems are critical to improving the efficiency of PAs while maintaining good signal quality. Sources of PA distortion and the impact of such distortions on the linearization system are briefly discussed, and an overview of the major categories of linearization techniques are presented with a focus on the predistortion system. The classification and comparison of various architectures of predistortion systems, based on their predistortion engine, is presented with a discussion of the predistortion function formulation. Various techniques to optimize the hardware implementation of DPD systems are also discussed, particularly recent efforts applicable to 5G wideband sub-6 GHz and mm-wave MIMO systems. The coefficient identification algorithm and recent developments aimed at reducing the speed requirement of the feedback path are also described.

Chapter 3 presents a novel digitally-assisted ARFPD (DA-ARFPD) architecture targeting sub-6 GHz small-cell base stations with reduced PA output power driven with wideband and CA communication signals. It presents a novel architecture with reduced hardware complexity and the corresponding linear small-signal-assisted parameter identification algorithm. A DA-ARFPD test bench incorporating major analog/RF components is described, and measurement results are presented which demonstrate its excellent linearization capability with gallium nitride (GaN) Doherty PA (DPA) driven by digitally modulated signals with a bandwidth up to 80 MHz.

Chapter 4 presents a novel real-time DPD hardware architecture for the linearization of 5G transmitters with ultra-wideband modulation signals as are often found in sub-6 GHz and mm-wave systems. To overcome the linearization bandwidth constraint imposed by the maximum clock frequency of the digital circuitry, a new parallel-processing DPD engine architecture is devised to allow multiple samples to be processed per clock cycle. The proposed real-time DPD architecture is implemented in a commercial field-programmable gate array (FPGA) that achieves a scalable linearization bandwidth of up to 2.4 GHz with 300 MHz core clock rate for the digital circuits.

Chapter 5 presents an under-sampling scheme for the feedback path that enables the update of the predistortion parameters using a low-speed ADCs. The proposed approach is found to have comparable linearization capability compared to a conventional full-rate based indirect learning DPD, even with a significantly under-sampled feedback signal. The proposed approach can be applied in both the DA-ARFPD and real-time DPD systems to significantly reduce the sampling rate of the feedback path, hence lowering the cost and power consumption of the feedback and identification components.

Finally, Chapter 6 summarizes the contributions of this work and discusses possible areas of future research work.

Chapter 2

Background and Literature on Power Amplifier Linearization

2.1 Introduction

To meet the demands of growing data traffic, current LTE-A and next generation 5G wireless networks have introduced various technologies such as small-cell base stations, high efficiency modulation schemes, massive MIMO, and ultra wideband mm-wave technologies. Such advancements greatly increase the total system capacity, but at the same time introduce stringent signal quality requirements on the RF front-ends. This translates into contradictory requirements in terms of linearity and efficiency for the RF PAs, which dominate the overall transmitter's power consumption and distortion, and necessitates the use of linearization techniques. DPD systems based on the principle of applying the pre-inverse of a PA's nonlinear characteristics to the baseband signal have been the most popular solutions to enable the linear and efficient transmission of wideband signals in macro base stations with signal bandwidths of tens of megahertz. Yet, the reduced output power level of small-cell base stations and the ultra-wide transmission bandwidths of mm-wave MIMO systems create new challenges to existing PA linearization solutions. This chapter reviews the contradictory requirements of efficiency and linearity for RF PAs, presents a review of PA linearization techniques, and discusses recent research efforts that have aimed to address some of the issues.

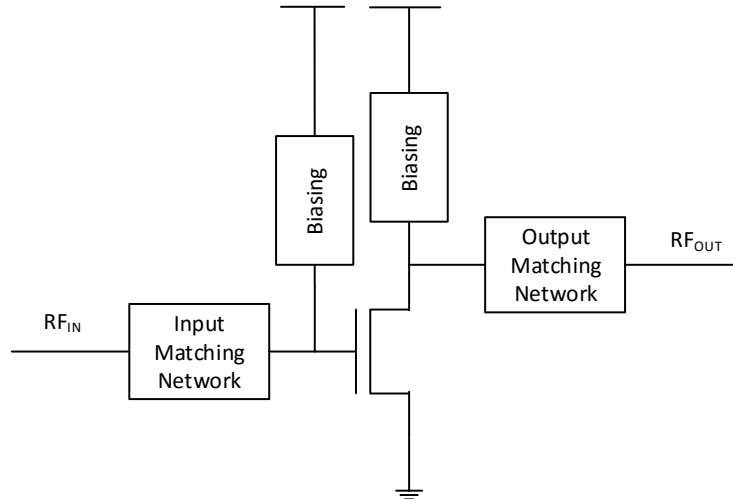


Figure 2.1: Block diagram of a conventional PA.

2.2 Power Amplifier Efficiency

PAs are power gain blocks that amplify a signal’s power before it is transmitted through the antenna. With large power output and high gain, a typical solid-state PA (as is commonly found in wireless communication base stations) is designed in an open loop fashion, as shown in the block diagram in Fig. 2.1. It consists of one or more transistors as active devices to perform the amplification, with biasing circuits and matching networks at the gate and drain of the transistors for the desired class of operation.

An ideal PA should provide linear amplification of the input signal with constant gain and phase shift across all frequencies and output power levels. However, a real PA is limited by device non-idealities and the constraints of the matching and biasing networks, leading to finite bandwidth, efficiency and linearity. Some of the key parameters of PAs are maximum output power, gain, efficiency and bandwidth.

Output Power The output power of a PA is characterized in terms of its saturation power P_{sat} —this defines the maximum output power of the PA. However, it is rare for a PA to operate close to P_{sat} as here the transistors reach saturation and the gain is heavily compressed, creating a lot of nonlinearity. A more useful specification of usable output power is P_{1dB} which represents the output power level at which the power gain is 1 dB smaller than the linear gain.

$$P_{1dB} = P_{out}|_{G==G_{linear}-1dB} \quad (2.1)$$

Gain Among the various definitions of amplifier gain, the power gain is the most relevant to a PA and is defined as the ratio of output power to input power.

$$G = \frac{P_{out}}{P_{in}} \quad (2.2)$$

The linear gain is typically measured using small signals at significant backoff power levels, i.e. power levels lower than P_{sat} , where the PA is linear. The large signal gain of a PA is expected to deviate from the small signal gain as the transistors approach saturation and the PA shows significant nonlinearity.

Efficiency The efficiency of a PA characterizes how well it converts the power drawn from the supply to useful output signal power. The power efficiency of a PA is measured as either the drain efficiency (DE) (2.3), defined as the ratio between the output power P_{out} and the total direct current (DC) power consumption P_{DC} , or as the power added efficiency (PAE) (2.4), which also takes the input signal power into account.

$$DE = (P_{out} - P_{in})/P_{DC} \quad (2.3)$$

$$PAE = P_{out}/P_{DC} \quad (2.4)$$

As long as the power gain of the PA is large, i.e. $P_{out} \gg P_{in}$, the DE and PAE are close.

Bandwidth The RF bandwidth of a PA defines the frequency range at which the PA can maintain its desired output power level. Wideband and multi-band PA designs include desirable features that allow one PA to cover multiple communication bands. Equally important is the achievable modulation bandwidth of a PA that describes the maximum bandwidth of the modulated communication signal that can be transmitted while maintaining good signal quality. This is closely related to the linearity of the PA and is important for wideband communication signals developed for modern wireless networks.

As the PA usually consumes the largest amount of power in the transmitter, its efficiency has a significant impact on the overall system's efficiency. Additionally, any DC power that is not converted into output power is released as heat, which will require additional cooling of the system and further degrade the overall efficiency. Hence, improving PA efficiency has always been a focus of PA development.

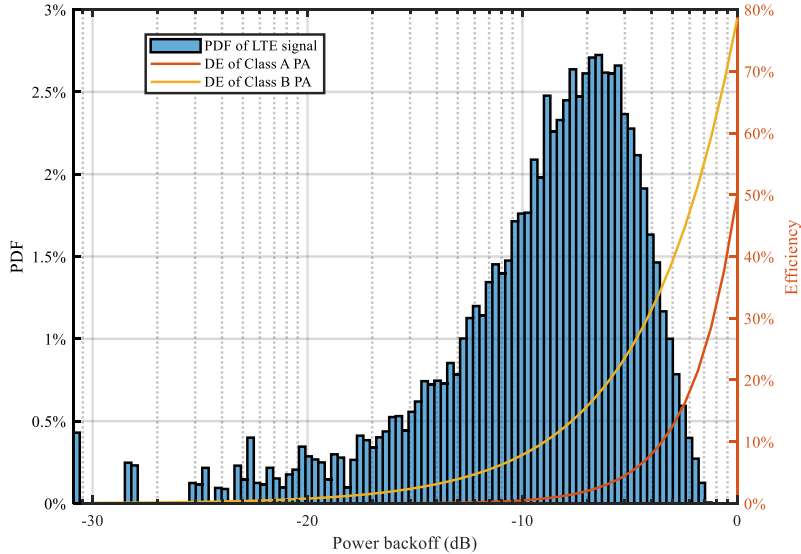


Figure 2.2: DE of ideal class A and class B PAs with the PDF of a LTE-A signal with 7.6 dB PAPR.

2.2.1 Conventional Power Amplifier Efficiency

Conventional PAs are often categorized into different classes of operation, ranging from class A, AB, B to C, based on how often the transistor is conducting, i.e., their conduction angle as set by the device biasing points[6, 7]. A class A PA is conducting all the time (conduction angle 2π), a class B PA is conducting half the time (conduction angle $= \pi$), and a class AB PA falls somewhere in-between. A class C PA is conducting less than half the time (conduction angle $< \pi$). The maximum attainable efficiency of a conventional PA occurs at the maximum output power level, with the theoretical maximum efficiency (using ideal transistors) increasing as the conduction angle decreases. The DE of an ideal PA changes from 50% in a class A PA, to 78.5% in a class B PA, and to approximately 100% for a class C PA with conduction angle approaching zero. A class AB PA falls in-between a class A and class B PA. However, this is at the cost of the power gain of the amplifier. In reality, device parasitics, losses in passive elements in the matching and biasing networks, as well as other nonidealities will further reduce the overall efficiency of the PA.

At the same time, the high efficiency modulation schemes used in today's communication systems result in a high PAPR of the modulated signal. This means the PAs are operating at significant backoff power levels. For conventional PAs, their efficiency de-

creases rapidly with output power level, resulting in further decreases in efficiency when transmitting such signals. This is illustrated in Fig. 2.2, where the DE of an ideal class A and an ideal class B PA are plotted together with the probability distribution function (PDF) of an LTE-A signal with PAPR of 7.6 dB. As the peak likelihood of the signal power level occurs at a significant backoff power level, a conventional PA is operating at a low efficiency level. The expected DE of an ideal class A PA would be less than 10% when transmitting this signal—meaning over 90% of the DC power is wasted and released as heat to the system. A class B PA would achieve significantly higher efficiency compared to a class A PA, with an ideal class B PA expected to achieve less than 40% DE when transmitting the same signal at the cost of lower gain. A class AB PA would have higher gain than a class B PA but lower efficiency. A class C PA could achieve higher efficiency, but it has the lowest maximum output power, gain and linearity among the conventional classes of PAs and is rarely used as the main PA of a transmitter. The mismatch between the power level of maximum PA efficiency and the maximum likelihood of high PAPR signals causes conventional PAs to have significantly lower average efficiencies compared to their peak efficiency. The PAPR of communication signals is going to increase further as higher level modulation schemes are used, further decreasing the efficiency of PAs.

2.2.2 Backoff Efficiency Enhancement Techniques

The poor efficiency of conventional PAs when transmitting modulated signals is due to the mismatch between the maximum likelihood of the signal power level and the peak efficiency of the PA. To address the aforementioned problem and improve the efficiency when transmitting modulated signals with high PAPR, advanced PA architectures that improve the backoff efficiency have been developed, such as drain modulation and load modulation.

Drain Modulation

Drain modulation is based on the principle that if the supply voltage at the drain of the transistor is lowered when the output power of the PA is lowered, the efficiency of the PA can be kept unchanged. To maintain high efficiency for a modulated signal where the signal output power is changing, the drain supply voltage needs to follow the signal, hence the name drain modulation. The ET solution is the most well-known approach to use supply modulation techniques [8, 9, 10, 11]. As illustrated in Fig. 2.3, the ET solution includes an envelope amplifier that modulates the drain supply voltage of the main PA according to the envelope of the signal. This lowers the DC power consumption when the signal power

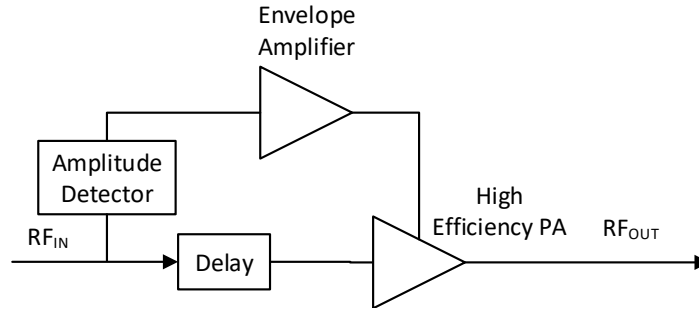


Figure 2.3: Block diagram of an ET PA.

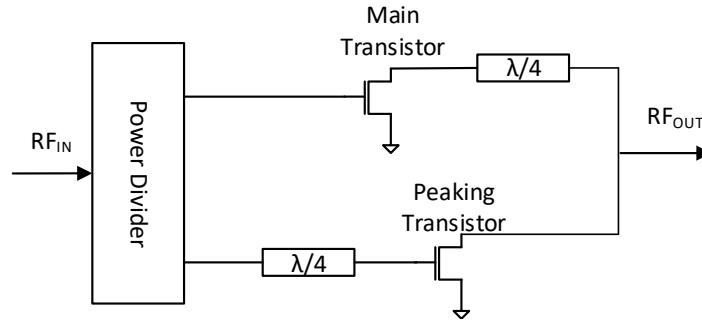


Figure 2.4: Block diagram of a DPA.

is small, thus improving the efficiency at backoff power levels. The ET solution has found wide adoption in handsets and other mobile devices as supply modulation can be readily integrated with supply regulation [11] resulting in low complexity solutions. While the RF bandwidth of ET follows that of the main PA, the main disadvantage of ET is the limited modulation bandwidth, primarily due to the limitation of the envelope amplifier that needs to modulate the drain supply voltage of the main PA. The envelope of a modulated signal is given by $\sqrt{I^2 + Q^2}$, where I and Q are the in-phase and quadrature (IQ) components of the signal, which has significantly wider bandwidth (3-4 times) compared to the complex modulated signal. Various techniques have been introduced to increase the modulation bandwidth [10, 12] by allowing the output of the envelope amplifier to deviate from the exact envelope of the signal in order to simplify the design of the envelope amplifier.

Load Modulation

While drain modulation lowers the supply voltage to reduce DC power consumption when the PA is operating at backoff power levels, the load modulation relies on increasing the effective load presented to the transistor at lower output levels. DPAs represent a family of load modulation techniques [13, 14, 15, 16]. As illustrated in Fig. 2.4, typical DPA designs bias the main and peaking transistors differently and provide efficiency enhancement at the 6 dB backoff power level. When input power is low, only the main transistor is conducting and the DPA acts as a conventional PA. When the power level is in the last 6 dB region, the peaking transistor starts conducting and modulates the load presented to the main transistor through the quarter-wave transmission line which acts as an impedance inverter. The effective load seen by the main transistor increases with increasing current from the peaking transistor as the output power increases, thus maintaining high efficiency operation. At peak power, both transistors are presented with the optimal load to achieve maximum efficiency. While a conventional DPA is designed to provide enhanced efficiency at the 6 dB backoff power level, other work has allowed for efficiency enhancement at further backoff power levels [13, 15]. Extensive research work on DPAs has demonstrated the great potential of this technique and DPAs have become one of the most widely used PA architectures in today's base stations [17].

2.2.3 Other Efficiency Enhancement Techniques

Both drain modulation and load modulation techniques focus on improving the overall efficiency of the PA when transmitting a modulated signal with significant PAPR by enhancing the backoff power efficiency. These techniques are well suited for integration with existing transmitters as only minimal changes to the transmitter's architecture are required. Yet, when more freedom is allowed, other efficiency enhancement techniques exist, such as outphasing, switched mode PAs, and digital PAs.

Switched Mode Power Amplifiers

Conventional PAs operate the transistor in the linear region as a voltage controlled current source that conducts current with respect to the input voltage. On the other hand, the transistor can be operated as a switch with only on and off states, resulting in a highly efficient switched mode PA. Examples of switched mode PAs are voltage/current mode class D PAs and class E PAs, with theoretical DEs of 100% [6, 7, 18, 19]. By operating the transistor as a switch, the transistor is either completely on with minimal resistance when

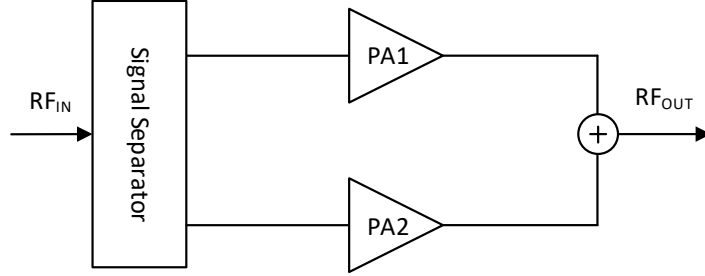


Figure 2.5: Block diagram of an outphasing PA.

conducting current, or completely off with no current flow. Hence the undesired power dissipation over the transistor is minimized, resulting in high efficiency. The main limiting factor of switched mode PAs is the transition frequency of the transistor, which must be high enough for the transistor to maintain near ideal switching behaviour at the operating RF frequency. With the advancement in semiconductor technologies, transistors with transition frequencies in the hundreds of gigahertz are now available, enabling switched mode PAs in the gigahertz range.

Outphasing

The outphasing technique aims to achieve high efficiency linear amplification through controlled combining of high efficiency nonlinear components, as illustrated in Fig. 2.5 [20, 21, 22]. The input signal, which can be expressed as

$$u_{in}(t) = A(t)\cos(\omega t + \phi(t)) \quad (2.5)$$

with $A(t)$ and $\phi(t)$ the magnitude and phase of the baseband, is split into two constant amplitude signals with different phases as follows

$$\begin{aligned} u_1(t) &= \cos(\omega t + \phi(t) + \cos^{-1}(A(t))) \\ u_2(t) &= \cos(\omega t + \phi(t) - \cos^{-1}(A(t))) \end{aligned} \quad (2.6)$$

and the combined output is

$$u_{out}(t) = G(u_1(t) + u_2(t)) = 2GA(t)\cos(\omega t + \phi(t)) \quad (2.7)$$

which is linear with respect to the input signal. Hence, outphasing is also known as linear amplification with nonlinear components (LINC). A combiner is the critical component in

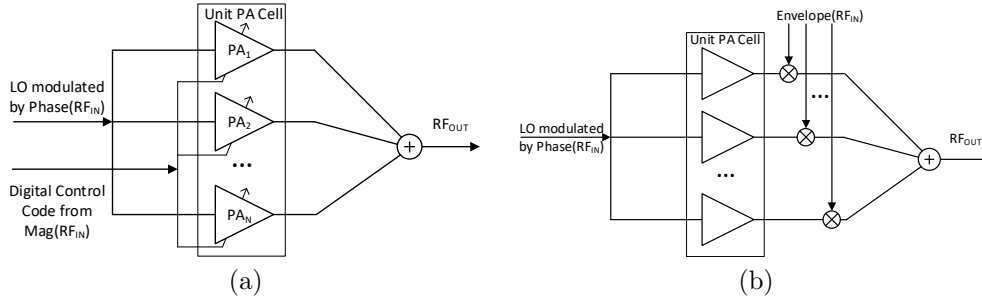


Figure 2.6: Block diagram of digital PA (a) and power mixer (b).

an outphasing system, and can be either isolating or non-isolating. An isolating combiner, like a Wilkinson power divider, provides a constant load to each PA at all signal power levels. This ensures good linearity of the outphasing system, but results in low efficiency at backoff power levels due to each PA still outputting at the maximum power level. On the other hand, non-isolating combiners such as the lossless Chireix combiner provide varying impedance to each PA based on the signal power level, hence improving the overall efficiency of the outphasing system. However, this comes at the cost of reduced linearity due to PA impedance mismatch at different power levels. Recent research work has further improved outphasing techniques by introducing mixed-mode operation and combining it with other efficiency improvement techniques [22, 23, 24].

Digital Power Amplifier and Power Digital-to-Analog Converter

A digital PA is the concept of turning on and off different numbers of unit PA cells from an array of such cells according to the magnitude of the signal [25, 26, 27]. As illustrated in Fig. 2.6a, the phase of the output is controlled by the phase modulated signal from the input signal that feeds into all unit cells, while the magnitude of the output is controlled by turning on the appropriate number of unit PAs based on the digital code word generated from the envelope of the input signal. Each unit cell ideally operates at either peak power, where the efficiency is highest, or completely switched off. This allows high efficiency architectures such as switched mode to be used for individual cell PAs while good efficiency can be maintained at all power levels. Together, their small area and ease of integration with other transceiver circuits make digital PAs attractive for many complimentary metal-oxide-semiconductor (CMOS) implementations. However, such architectures suffer from the inherent tradeoff between the speed and power of semiconductor devices, and have limited bandwidths and power dynamic ranges, as well as unwanted quantization noise.

An improvement to the digital PA concept is to use a mixer as both the frequency conversion and power generation elements [28, 29, 30, 31]. An example of such a scheme is a power mixer array, as illustrated in Fig. 2.6b. The phase modulated local oscillator is fed to all unit mixer cells, and mixed with the baseband envelope. Commonly using a segmented power generation scheme [28], each power generating mixer is switched on based on the envelope of the input signal. Using the mixer as the power generation unit brings the benefits of reduced noise and larger power dynamic range. However, it does come with increased area size and higher complexity.

2.3 Power Amplifier Linearity

High efficiency modulation schemes such as QAM used in modern communication networks require both amplitude and phase modulation of the carrier. The constellation diagram of a 16-QAM modulation scheme is shown in Fig. 2.7a. Each symbol has a different magnitude and phase—when combined with a higher symbol rate the resulting modulated signal exhibits both amplitude and phase variation over a wide bandwidth. There are many sources of distortion along the communication link, such as noise, interference, channel dispersion, etc., that cause the received symbol to deviate from its intended position, as illustrated in Fig. 2.7b. Among all sources of error, the PA nonlinearity is usually the most dominant on the transmitter side. Under this type of modulation scheme, the error of the received signal needs to be kept below a certain threshold to ensure the correct identification of the symbols. With increased spectral efficiency comes denser constellations with smaller distances between the constellation points, thus less error can be tolerated when attempting to maintain a low bit error rate. This translates into stringent requirements for RF PAs to maintain good linearity at wide output power ranges and bandwidths.

The linearity of a PA measures the difference between its output signal to a linearly scaled version of its input signal. Due to limitations and imperfections of transistors and other components, a real PA never achieves the desired perfect linear amplification of the signal. Any deviation of the PA’s output from the desired signal, except a constant gain and phase shift, is viewed as distortion. From the PA linearization perspective, the effects of PA distortion are usually expressed in terms of error vector magnitude (EVM) or normalized mean square error (NMSE) that measures how closely matched the output signal is to the desired signal. EVM measures the deviation of the actual symbol from the ideal constellation position as shown in Fig. 2.7 and is defined as

$$\text{EVM}(\%) = 100\% \sqrt{P_{\text{error}}/P_{\text{ref}}}. \quad (2.8)$$

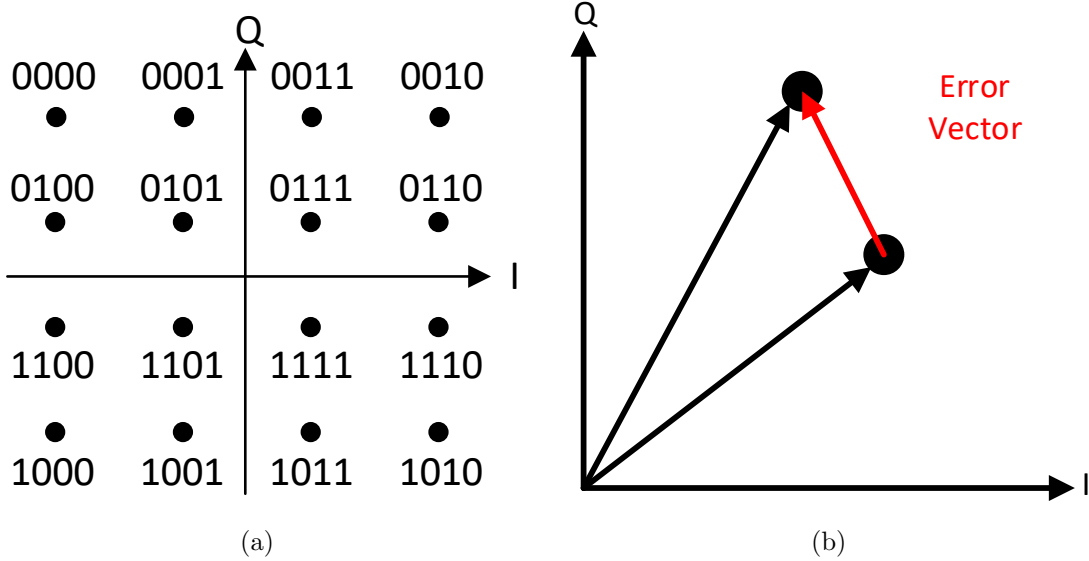


Figure 2.7: Constellation of 16-QAM (a) and error vector (b).

where P_{error} and P_{ref} are the root-mean-square power of the error vector and ideal signal vector, respectively. NMSE measures the mean square error between the sampled output signals with respect to the sampled input signals, and is defined as

$$NMSE = \frac{1}{N} \frac{\sum_i (x_i - y_i)^2}{\sum_i x_i^2} \quad (2.9)$$

where x_i and y_i are the i th samples of the desired and actual signals.

Besides causing deviation in the transmitted symbols, PA distortion also generates additional frequency components in the output spectrum due to the nonlinear response. The most important type of distortion is adjacent band distortion; it is difficult to filter out in practice. It has the effect of spreading the signal power outside its assigned frequency channel, thus it leaks into adjacent channels, hence the name spectrum regrowth. Such leakage acts as unwanted interference to the users of the adjacent channels and degrades their signal to interference and noise (SNIR) ratio. Therefore, it needs to be minimized. This effect is measured in terms of ACLR that is defined as the ratio of the transmitted RF power on the assigned channel to the power in the adjacent channel, often expressed in decibels (dB).

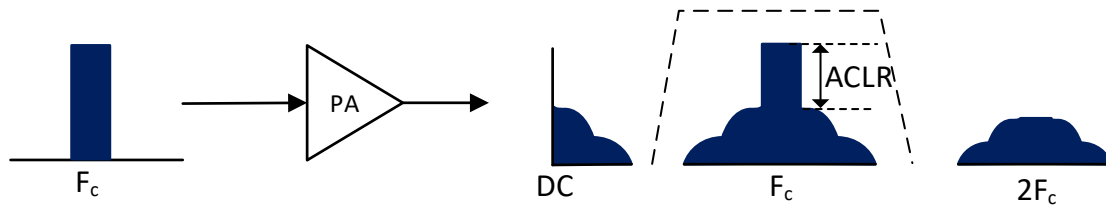


Figure 2.8: Illustration of PA spectra regrowth.

2.3.1 Sources of Power Amplifier Distortion

PA distortion can come from various sources, ranging from the transistor itself, the biasing and matching networks, or a complex system level design. Some significant sources of PA distortion will be briefly discussed in the following sections.

At the individual PA level, non-idealities in the transistor and biasing and matching networks are the main contributors to the distortion.

Nonlinear Transconductance Transconductance of a real transistor is not constant over all voltage ranges. Effects such as a soft turn-on where the drain current flowing through the transistor increases gradually when the gate voltage is just above the threshold voltage of the transistor, and gradual saturation at maximum drain current can be observed. This results in variations in gain and phase that cause nonlinearity over different input powers [32].

Knee Voltage Unlike in an ideal transistor model, there are no abrupt transitions from the linear region to the saturation region in a real transistor. When the drain voltage swings into the knee region, the drain current decreases rapidly and in turn reduces the gain of the PA, causing static nonlinearity.

Nonlinear Capacitance The parasitic capacitance at the gate, source and drain of a real transistor vary nonlinearly versus voltage. As they are non-negligible in a typical RF PA, the varying capacitance contributes to magnitude and phase variations of the output, thus causing static nonlinearity [10].

Non-ideal Biasing Networks A realistic biasing network has a finite impedance and limited isolation. This in turn causes feedback from the drain to the gate, particularly at the baseband frequency range. Such unwanted feedback modulates the input and interacts with other distortion mechanisms to generate nonlinearity in the PA.

Non-ideal Matching Networks A realistic matching network consists of transmission lines and lumped elements that have non-flat frequency responses across the PA operation bandwidth. This causes a changing impedance presented to the PA across the signal bandwidth, and in turn causes variations in the gain, phase and group delay of the PA across the signal bandwidth. Such distortion becomes more significant with wider signal modulation bandwidths [33].

Thermal and Trapping Effects As PA output powers vary when transmitting modulated signals, the heat emission and temperature of the transistor also change over time. This temperature variation changes the transconductance of the transistor, creating dynamic nonlinearity that depends on the signal. In addition, PAs using III-V device transistors suffer from trapping effects that effect the threshold and other characteristics of the transistor, introducing addition distortion [34].

The backoff efficiency enhancement techniques discussed in the previous section introduce additional sources of distortion to the PA, as more complex design topologies further degrade the linearity.

Envelope Amplifier Nonlinearity The envelope amplifier in an ET system can only present a finite impedance to the drain of the main transistor. This causes undesired drain modulation among other effects. The nonlinearity in an envelope amplifier has been found to introduce significant distortion into the ET system's output [10, 35]. Time alignment mismatch between the envelope and the main signal path also contribute to the distortion of an ET system.

Doherty Power Amplifier Nonlinearity The quarter-wave transformer in a DPA only acts as an ideal impedance converter at the center frequency; for wideband signals it introduces magnitude and phase responses across the signal bandwidth [36]. Non-idealities in the peaking transistor cause distortion through the load modulation mechanism, resulting in a nonlinear response in the DPA output.

The other efficiency enhancement techniques discussed such as outphasing and digital PAs also have their own inherent sources of distortion arising from their design.

On top of the distortion of individual PAs, in massive MIMO systems, the closely packed PAs and antennas create new sources of distortion. Cross-talk occurring before the PAs due to limited isolation of the signal path and biasing networks, results in PAs on different paths interacting with each other and creates complicated nonlinear distortion. The closely spaced antennas in the array create non-negligible mutual coupling, and generate varying loading to the PAs that depends on the location of the antenna and array settings [37].

2.3.2 Modelling of Power Amplifier Nonlinearity

Since the linearity of an RF PA is important to achieving a high data rate, it is necessary to have a good model of the PA's nonlinearity in order to predict distortion effects on signal quality and to provide guidance to improve its linearity. As discussed in the previous sub-section, RF PA distortion can come from many sources and the combination and interaction between various sources of distortion result in complex dynamic nonlinear behaviour. It is difficult to describe such complex transfer functions using physical models (like compact models) which are commonly used to describe devices and circuits. Furthermore, the computational complexity required to simulate using such a model is high. Thus, PA nonlinearity modelling typically uses behaviour models that describe the relationship between the input and output signals for a particular PA under certain operating points.

From a system perspective, the distortion of a PA can be described as a mapping between the input signal and the normalized output signal. In a discrete system with sampled input and output, the PA output can be described as

$$y[n] = g_n x[n] \quad (2.10)$$

where $x[n]$ and $y[n]$ are the n th normalized samples of the PA's input and output signals, and g_n is the complex gain of the respective sample. An ideal PA has a constant gain of unity for all samples, where distortion of the PA causes deviation in magnitude and phase. The effects of PA distortion on the transmitted signal can be visualized as amplitude-to-amplitude (AM-AM) and amplitude-to-phase (AM-PM) responses as shown in Fig. 2.9, where the gain and phase of the samples of normalized output with respect to the input sample are plotted against the input power. Taking a closer look at the AM-AM and AM-PM figure, it is clear that for samples with the same input power, the output can have a variation of gain and phase distortion. This scattering in AM-AM and AM-PM responses indicates that the PA output is not only a function of the current input but the past as well.

As a result, the PA distortion can be categorized into static and dynamic distortion.

Static Distortion This distortion represents the deviations in the gain and phase corresponding to increased input power that are dependent solely on the current input signal, shown as the red trace in Fig. 2.9. The static distortion function has a one-to-one mapping for each input power level, and is independent of the signal characteristics. It can be characterized by performing a power sweep using a continuous wave (CW) signal and computing the gain and phase of the output and comparing them to the desired gain and phase. Static distortion is a good approximation of the PA's nonlinearity using narrow band signals.

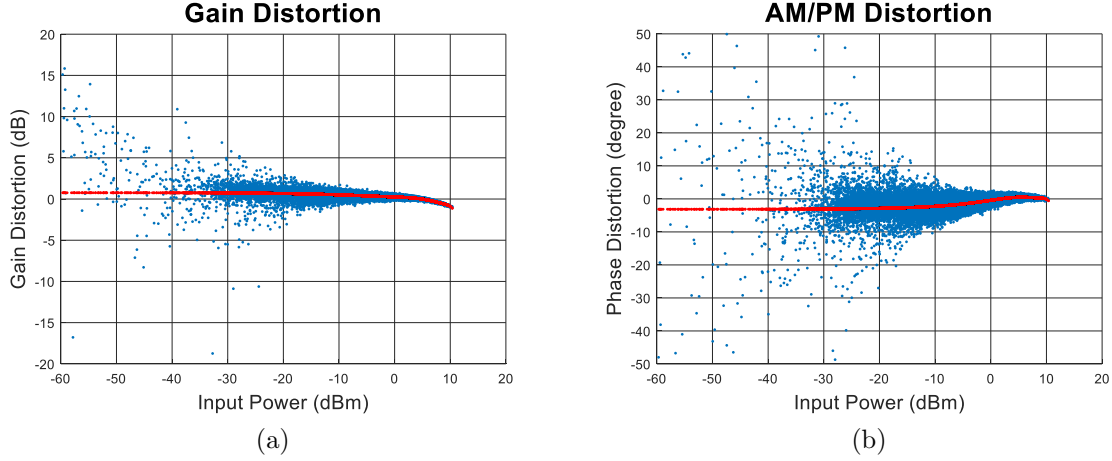


Figure 2.9: Gain (a) and phase (b) distortion of PA under modulated signal, with static distortion highlighted.

Dynamic Distortion This type of distortion, on the other hand, represents the deviations in the gain and phase that depend on both the current and past input signals. It describes the distortion behaviour in which the previously transmitted signal has a lasting effect on current and future transmitted signals. The distortion becomes signal dependent and changes dynamically, as reflected in the scattering of the gain and phase in Fig. 2.9. Hence, it is often referred to as memory effects, as if the PA remembers what has been transmitted previously [38]. The dynamic distortion can be further divided into linear and nonlinear memory effects, depending on the relationship with the input signals. Linear memory effects represent the distortion associated with the linear elements in the PA, such as the frequency response of the matching network. Nonlinear memory effects are the results of interactions between the nonlinear components and the rest of the circuitry, creating the most complex dynamic behaviour in the PA. As signal bandwidth increases, the memory effects of the PA also increase significantly.

To create a behaviour model for a PA, certain assumptions are made based on our understanding of practical PAs used in wireless communication systems. The PA is treated as a weakly nonlinear device in the sense that nonlinearity only acts as perturbation around the operation point. The PA is also assumed to have a fading memory, such that the impact of the current signal on later signals diminishes with time. These assumptions fit naturally into a Volterra series model as in (2.11), which is a combination of linear and nonlinear

multidimensional convolution [39].

$$y(t) = \sum_{i=1}^N \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_p(\tau_1, \dots, \tau_p) \prod_{j=1}^i x(t - \tau_j) d\tau \quad (2.11)$$

where $x(t)$ and $y(t)$ are the input and output RF signals respectively, $h_p(\tau_1, \dots, \tau_p)$ denotes the Volterra kernels with order p , and N is the nonlinearity order. The baseband equivalent and discrete form of (2.11) is

$$\tilde{y}[n] = \sum_{i=1}^N \sum_{m_1=0}^M \dots \sum_{m_N=0}^M \prod_{i=1}^i h_i(m_1, \dots, m_i) \tilde{x}[n - m_i] \quad (2.12)$$

where $\tilde{x}[n]$ and $\tilde{y}[n]$ are the discrete baseband equivalent envelope samples of the input and output signals respectively, N is the nonlinearity order, and M is the memory depth. This polynomial form has all the aforementioned desirable features, and has become a popular choice for modelling PA distortions.

One major disadvantage of the Volterra series formulation is that the number of kernels increases exponentially with nonlinearity order and memory depth. This makes it impractical to use in real-time applications due to over-modelling and computational complexity. The large number of kernels also leads to numerical instability, particularly during the least squares (LSs) identification process. This has led to major research efforts to reduce the complexity of the Volterra series and identify the dominant kernels [40, 41, 42, 43, 44]. A large part of such work has been aimed at the field of DPD, one of the most popular forms of linearization techniques, and it will be discussed in later sections.

Another family of PA behaviour models uses artificial neural network (ANN) to capture the PA distortion, such as [45, 46, 47]. ANN-based models are valuable due to their excellent capability to accurately approximate nonlinear functions, as the universal approximation theorem [48] has proven that a feed-forward ANN with at least one hidden layer and non-constant activation function can approximate arbitrary nonlinear functions with any desired error. The main research focus of ANN-based PA models is to choose the correct network size and training method. Once trained, the feed-forward ANN is a fast and efficient model to predict the output of PAs under nonlinear distortion.

2.3.3 Linearity versus Efficiency Trade-off

Linearity and efficiency are contradictory goals in RF PA design. For a typical PA, linearity is good at low output power levels, but efficiency is poor. On the other hand, efficiency is

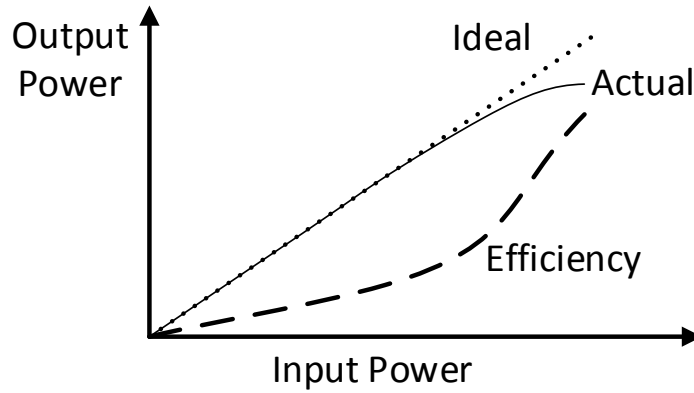


Figure 2.10: Linearity versus efficiency of PA.

high at large output power levels, but the linearity of the PA suffers. Such linearity and efficiency trade-offs are illustrated in Fig. 2.10. It is a delicate balance to find the optimal operation point where a PA can achieve maximum efficiency without compromising signal quality due to degradation in linearity.

As wireless communication networks have evolved to meet the demand of growing data traffic, the linearity requirement for RF PAs has become more stringent due to the complex modulation schemes and advanced access technology. At the same time, efforts to achieve higher efficiency have pushed the PA towards operation regions with more pronounced nonlinearity. The use of advanced PA topologies and techniques introduces additional distortion mechanisms and the widening of signal bandwidths further complicates the situation. Thus, an approach emerged where the linearity-efficiency tradeoff was no longer addressed only at the RF PA, but also at the system level through the use of linearization techniques. Linearization systems have seen rapid development and wide adoption in today's wireless communication systems, playing a vital role in the transmitter system's ability to achieve high efficiency and linearity.

2.4 Linearization Techniques

As per the discussion in previous sections, RF PAs face increasing challenges in meeting both the contradictory requirements of efficiency and linearity. Linearization systems aim to relax the constraints on the PA by providing a system-level solution to counteract the distortion introduced, thus enabling the PA to operate in the high efficiency region and use

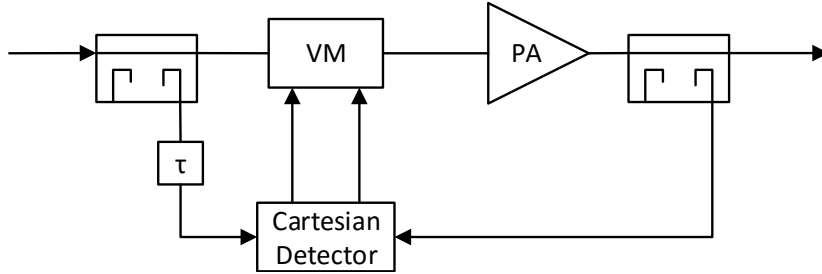


Figure 2.11: Block diagram of a feedback linearization technique

advanced architecture. Many PA linearization techniques such as feedback, feedforward and predistortion have been developed [49]. Over the years, DPD systems, being highly effective, inherently adaptive and readily integrable with existing transmitter architectures, have become the predominant linearization systems in use. This section provides a review of these linearization techniques, starting with an overview of different families of linearization techniques, followed by a detailed review of the DPD system.

2.4.1 Feedback Techniques

The feedback technique has been widely used in analog circuits to stabilize the gain of an amplifier. It is also used in RF PAs where the output is constantly tracked and corrected through a negative feedback mechanism, such that the gain of the PA is maintained unchanged despite changes in output power [50]. Following classical feedback theory, the gain of the amplifier with feedback is set by

$$\frac{v_o}{v_i} = \frac{A_o}{1 + \beta A_o} \quad (2.13)$$

where A_o is the open loop gain of the amplifier and β is the feedback factor. The larger the feedback loop gain $1 + \beta A_o$, the more stable the closed loop gain A is with respect to the open loop gain A_o of the amplifier—hence, there is greater linearity. However, this comes at the price of lower gain as A decreases with larger feedback loop gains.

For a PA operating at an RF frequency, the open loop gain A_o is not very high, and a large feedback factor β is often avoided due to stability concerns. Hence, this direct feedback form is rarely used in practice. The indirect feedback technique as shown in Fig. 2.11 compares the envelopes of the output and input signals, and the difference is

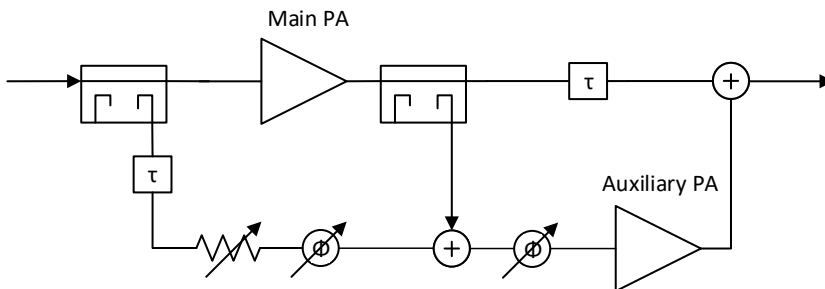


Figure 2.12: Block diagram of a feedforward linearization technique.

used to modify the input to reduce the distortion. Alternative solutions using Cartesian feedback, which uses the in-phase and quadrature components of the signal, have also been used. This approach removes the need for a phase shifter. Another major disadvantage of the feedback technique is its limited bandwidth, as it is difficult to maintain the required phase of the feedback signal (180°) across a wide frequency band for the feedback circuits. The benefits of negative feedback quickly diminish as the phase of the feedback signal deviates from the optimal; the approach can even lead to instability if the feedback signal phase results in positive feedback.

2.4.2 Feedforward Techniques

The feedforward linearization technique is built around the principle of active distortion cancellation [51]. The signal is amplified in the main PA and the distortion of the output is then feedforward with 180° phase shift and combined destructively with the output from the PA. The typical configuration of a feedforward system is illustrated in Fig. 2.12. The input and output of the main PA are sensed through a set of couplers and compared. The comparison is done by first matching the sensed input and output signals in magnitude but with opposite phase through a series of attenuators and phase shifters. The signals are then combined to extract the error signal containing the distortion information. The error signal obtained in this way is then amplified linearly through an auxiliary amplifier to match the magnitude of the distortion in the main path. The phase and group delay of the two paths are carefully adjusted through delay elements and phase shifters to ensure a destructive combining at the output of the system. In this way, the distortion of the main PA is cancelled, leaving only the desired linear component to be transmitted.

Feedforward linearization systems have very good distortion cancellation over a wide bandwidth, with over 20 dB improvement in ACLR reported [52]. However, the main

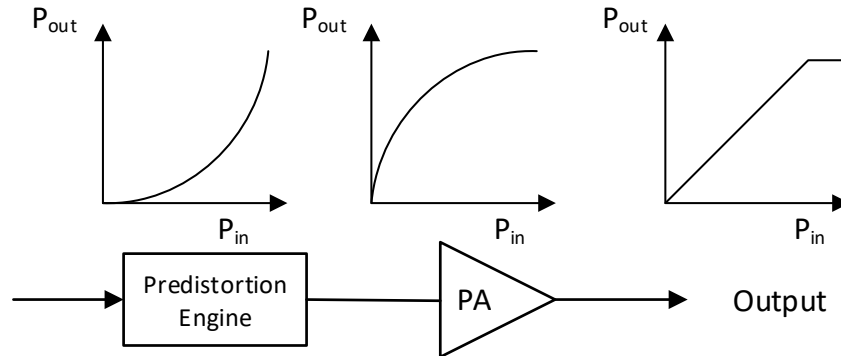


Figure 2.13: The principle of predistortion linearization techniques.

disadvantage of the feedforward technique is its high power overhead that leads to low overall efficiency of the system, limiting the benefits of using it for PA linearization. This is mainly due to the power consumption of the auxiliary amplifier, which needs to maintain high linearity in order to accurately amplify the error signal. Such an auxiliary amplifier is usually biased in class A and operates in deep backoff, leading to very low efficiency and high power consumption. The power consumption of the auxiliary amplifier increases as the distortion in the main PA worsens. Combining at the output of the main PA also incurs non-negligible insertion losses. The losses due to the couplers, phase shifters and combining networks further decrease the overall efficiency of the system. The complexity of implementation and difficulties integrating with existing amplifiers, together with the power overhead, limit the adoption of feedforward linearization techniques for RF PAs used in wireless communication networks.

2.4.3 Predistortion Techniques

Contrary to cancelling the distortion afterwards as in the feedforward approach, predistortion techniques seek to engineer the input signal fed to the PA to counteract the distortion effects such that the output is linear with respect to the original signal [50]. As illustrated in Fig. 2.13, the input signal is first processed by a predistortion engine, which itself has a nonlinear transfer function that is the inverse of the PA distortion. The output signal from the predistortion engine is such that, when amplified through the PA and subject to its distortion, it results in a combined transfer function that is linear.

The desired nonlinear function can be realized in various ways. Early work in the predistortion linearization field relied on the unique nonlinear characteristics of analog and

RF devices, such as Schottky diodes that have an exponential response. Tuning can be done through the sizing and biasing of the device, and a more accurate approximation of the desired nonlinear function can be achieved by using multiple devices. Memory effects can also be modelled through the use of multiple branches with delay taps. As progress in wireless communication technologies imposes more stringent linearity requirements and wider signal bandwidths, DPD has become more popular, taking advantage of the rapid increase in computation power offered by the digital signal processors (DSPs) in nano-scale process technologies. The inverse functions are typically modelled as families of polynomial functions of the Volterra series due to the ease of implementation. Coefficients for these models are founded based on the theoretical framework of p th-order inverses of nonlinear systems which has shown that the pre-inverse of a nonlinear system is identical to its post-inverse [53]. ANN-based models, used for PA behaviour modelling, can also be used as DPD engines in a similar fashion.

DPD has benefited from rapid advancements in digital integrated circuit technology, as its power consumption has been reduced with each new generation of process technology node. Its digital nature means its performance has good immunity to analog distortion in the components, and also makes it readily integrable with the baseband processor. Thus, DPD has been widely adopted in industry and is the dominant form of linearization solution used in wireless communication networks currently.

2.5 Literature Review on Predistortion Techniques

With a focus on mobile data, PA linearization techniques—in particular DPD—have attracted increasing attention from both academic and industrial researchers. Aiming to improve the linearization capability, reduce power and cost overhead, as well as address new distortion mechanisms introduced by advanced PA or MIMO systems, a wide range of research objectives have been pursued, ranging from novel architectures for the predistortion system, new formulations of the predistortion engine, advanced feedback and training methods, and implementation optimization of the predistortion system. The rest of this section presents an overview of the predistortion system, followed by a review of recent research efforts in the abovementioned areas as researchers continue to improve the predistortion system to address emerging requirements imposed by advancements in wireless communication systems.

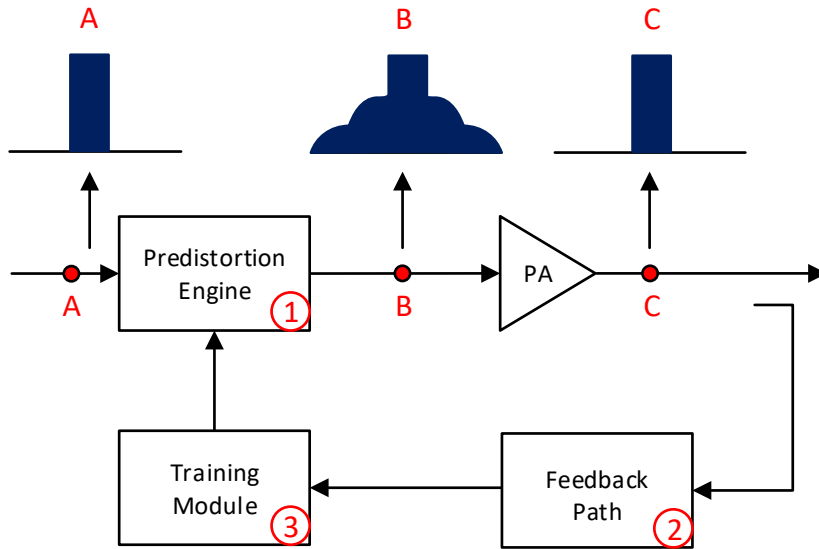


Figure 2.14: Block diagram of a predistortion system

2.5.1 Overview of Digital Predistortion Techniques

DPD systems have become the most popular linearization approach thanks to their excellent linearization capability, ability to track changing PA conditions, ease of implementation, and ability to be integrated with baseband systems. A typical DPD system, as illustrated in Fig. 2.14, consists of three main building blocks:

1. the predistortion engine that applies the predistortion function
2. the feedback path that monitors the PA output
3. the training module that identifies the parameters of the predistortion engine

Predistortion Engine

The predistortion engine is responsible for applying the desired predistortion function to the input signal. As an inverse of the PA nonlinearity, the predistortion function is typically selected from families of polynomial functions from the Volterra series or as a particular realization of ANN. In most DPD implementations, the predistortion engine uses a fixed set of basis or kernels but allows the coefficients to be updated to adapt to changing PA

conditions. The selection of the basis or kernels has attracted active research interest as it determines the linearization capability of the DPD system.

Due to the nonlinear nature of the predistortion function, the output of the predistortion engine experiences spectrum expansion similar to that experienced by a nonlinear PA, as illustrated in Fig. 2.14. Thus, the input and output of the predistortion engine need to be oversampled to capture the entire expanded bandwidth—typically 5 times the signal bandwidth in most practical applications. As the predistortion engine is on the transmission data path, it needs to compute the results in real time to keep up with the data throughput. This, combined with the oversampling, translates into a large amount of computation at high speed, making the predistortion engine the most critical component in the predistortion system as it needs to be operating all the time and dominates the power consumption. As the output of the predistortion engine is converted into an analog signal using a set of digital-to-analog converters (DACs), the high sampling rate required by the predistorted signal means the DACs also need to support a high data rate, resulting in additional cost and power consumption. Hence, the efficient implementation of a predistortion engine in terms of optimized hardware resource usage and power is also an important area under research.

Feedback Path

The feedback path is responsible for monitoring the output of the PA and capturing the output signals for the training module to update the coefficients used in the predistortion engine. In a typical predistortion system, the feedback path follows the standard receiver design, with the PA output attenuated, then down-converted and filtered before being sampled by an ADC. Hence, the feedback path is also referred to as the TOR.

This TOR has more stringent requirements compared to the receiver used in the transmission system, particularly in terms of linearity and bandwidth. Very high linearity is required of the TOR, because any distortion in the TOR will be mixed with the PA distortion and cannot be distinguished by the training algorithm, thus degrading the linearization performance of the predistortion system. Large bandwidth is also required, since conventional designs require the TOR to capture the full spectrum of the in-band distortion that is typically 5 times the signal bandwidth. As the signal bandwidth increases, these requirements become more demanding, and wideband down-converters and high speed ADCs are needed. The frequency dispersion of the receiver across the signal bandwidth, as well as the cost and power consumption of the feedback path, would become a concern at larger bandwidths.

Training Module

The training module is responsible for identifying the coefficients used in the predistortion engine to linearize the PA. Training can be performed offline to provide an initial guess for the predistortion, assuming a particular operation condition. However, PA distortion can change over time due to changes in operation condition such as temperature and supply voltage as well as wearing of the device, and the initial set of coefficients may become suboptimal. Thus, a training module becomes an essential part of the DPD system for real-time online updates of the coefficients in response to changes in PA nonlinearity to form a close-loop adaptive linearization system. The training module is turned on to update the engine coefficients only if there is a need to do so due to changes in PA operating conditions. The convergence speed and computational complexity of the training algorithm are of major concern, with the power consumption coming as a secondary consideration.

Typically implemented in software, the training module is closely coupled with the formulation of the predistortion engine, as well as the feedback path. For predistortion models with linear-in-parameter, such as those families of polynomials based on Volterra series, conventional LSs estimation can be used. For other families of predistortion models that do not have such a linear relationship between the output and the model coefficients, such as the Wiener model or ANN, more complicated search and optimization techniques have to be used. The LSs estimation has good convergence characteristics and fast computation and is favoured compared to other methods.

2.5.2 Architectures of Predistortion Engine

While the DPD system implements the predistortion engine in the digital domain and applies it to the baseband signal, there are many other forms of predistortion besides the popular DPD system. As illustrated in Fig. 2.15, the predistortion function can be synthesized in the digital, analog or RF domains and can be applied to the signal in baseband or RF. Depending on whether the nonlinear function is synthesized in digital or analog and how the predistortion function is applied to the signal, the predistortion technique can be categorized into one of five different kinds as summarized in Table 2.1:

(1) Digital Predistortion

In a DPD system, the nonlinear function is synthesized in the digital domain and the predistorted signals are generated in the digital domain as well. Benefiting from the increased

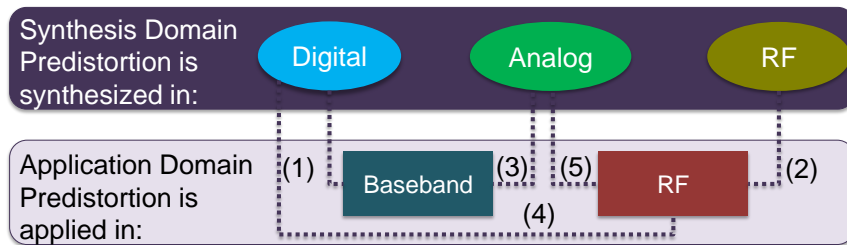


Figure 2.15: Classification of predistortion engines.

Table 2.1: Summary of the different classes of predistortion engine

	Class of predistortion	Synthesize Domain	Application Domain
(1)	Digital Predistortion (DPD)	Digital	Baseband
(2)	RF Predistortion (RFPD)	Analog	RF
(3)	Analog Predistortion (APD)	Analog	Baseband
(4)	Digital/RF Predistortion (DRFPD)	Digital	RF
(5)	Analog/RF Predistortion (ARFPD)	Analog	RF

processing power of DSP devices, thanks to rapid technology scaling, the DPD system has become the most popular predistortion system because of its power, linearization capabilities, and the ease of implementation. It has evolved from a simple lookup-table-based implementation [54] to sophisticated Volterra-series-based polynomial or ANN-based engines capable of handling wideband and multi-band signals [40, 55, 56, 44, 57].

Recent developments in 5G wireless technology have created additional challenges for the predistortion engine. The large number of PAs in the antenna arrays for mm-wave systems challenge the original concept of one predistortion engine per PA, requiring further innovation in the modelling. On the other hand, with significant increases in signal bandwidth, the data throughput needed by the predistortion engine can exceed the maximum clock frequency of digital circuits, calling for new implementation strategies. In addition, the power overhead of DPD increases with the signal bandwidth and does not scale down with the PA output level. This becomes a significant factor for small-cell base stations with much lower transmitted powers, making DPD less attractive compared to other more efficient methods.

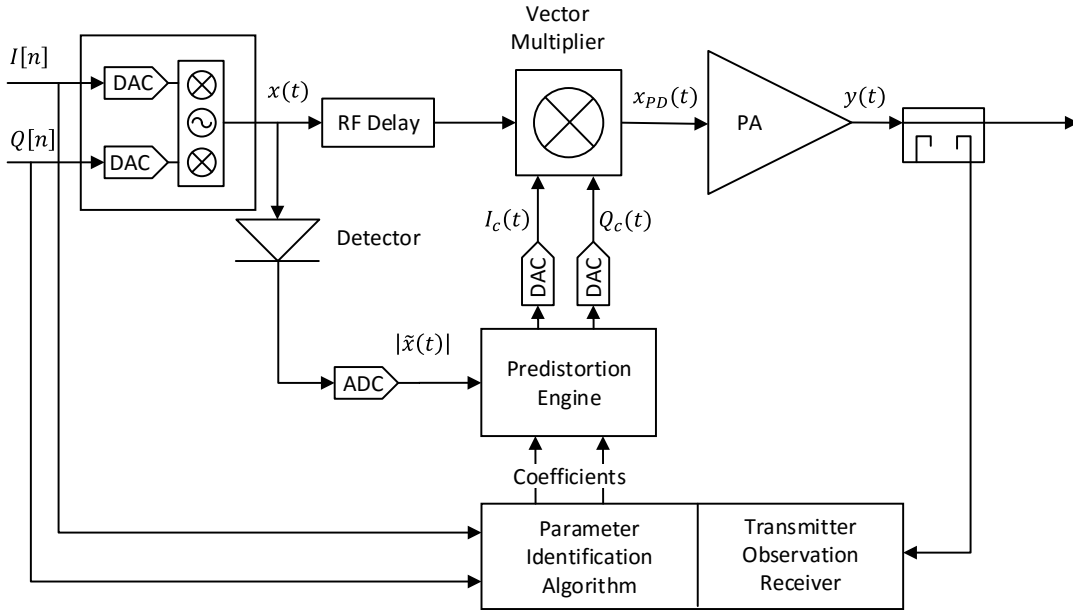


Figure 2.16: Block diagram of a digital/RF predistortion system.

(2) RF Predistortion and (3) Analog Predistortion

In an RF predistortion system, both the nonlinear function and application is in RF. In the analog predistortion engine, the nonlinear function is synthesized in analog and the predistorted signals are generated in analog as well. These full analog solutions achieve really low power consumption by removing all digital circuits [58, 59, 60, 61]. Static polynomials with nonlinearity orders of up to 5 are synthesized using analog circuits and applied directly to the intermediate frequency (IF) or RF signals. While being truly low-powered solutions, their practical adoption has been hindered by their limited linearization capacity in regards to static nonlinearity, and the lack of a viable solution to address memory effects.

(4) Digital/RF Predistortion

In a digital/RF predistortion scheme, the nonlinear function is synthesized in the digital domain and is applied to the RF signal through a vector multiplier. [62, 63, 64]. It accepts an RF input and generates a predistorted RF output, which is advantageous in situations where access to the digital baseband signal is not possible. However, the predistortion

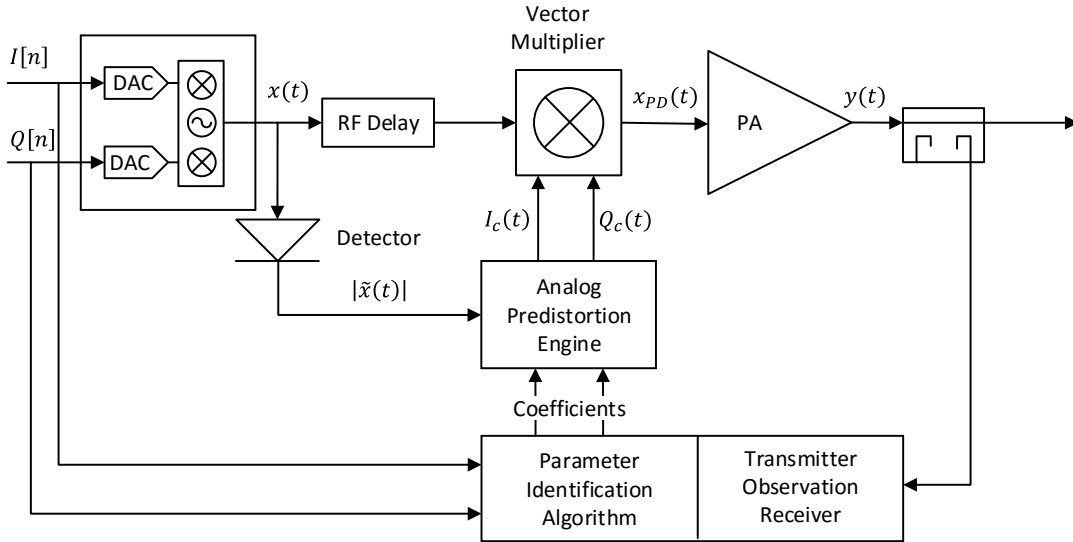


Figure 2.17: Block diagram of an analog/RF predistortion system.

engine is still implemented in the digital domain, which means it faces the same problem as DPD. The block diagram of a DRFPD system is shown in Fig. 2.16.

(5) Analog/RF Predistortion

In an analog/RF predistortion scheme, the nonlinear function is synthesized in analog while the predistortion is happening in RF [65, 66]. Similar to the DRFPD concept, the ARFPD engine generates the predistortion signal using a low-power analog engine that implements a polynomial predistortion function capable of handling memory effects. The vector multipliers then apply the predistortion function to the RF signal. The ARFPD engine eliminates the power hungry DPD engine and reduce the bandwidth requirement of the DAC, at the cost of additional analog hardware such as the analog predistortion engine and the RF vector multipliers. It retains the flexibility and performance of a DPD engine while leveraging the low-power nature of analog circuits. This framework has been investigated recently in an ARFPD system capable of linearizing PAs with up to 20 MHz modulated signals while consuming as little as 0.2 W of power has been reported [66]. The linearization capacity of the aforementioned work is still limited, however, due to the difficulty of designing a predistortion model of comparable linearization capacity to DPD using analog hardware. A block diagram of an ARFPD system is shown in Fig. 2.17.

2.5.3 Predistortion Function Formulation

Regardless of the class of predistortion engine, the formulation of the predistortion function is critical to the linearization capability. As the inverse of the PA distortion, the predistortion function is also a weakly nonlinear function with fading memory. It is not surprising that many of the behaviour models in Sec. 2.3.2 for the PA can be used as predistortion functions.

While there are many types of behaviour models for nonlinear systems, for the purpose of PA linearization there are certain features that are desirable in the PA distortion model:

- The model should be in the discrete time domain as communication signals are often sampled and processed in digital circuits.
- The model should have good modelling capacity and handle both static nonlinearity and memory effects.
- The model should be robust and able to fit different types of PAs under various operation conditions.
- The model should work on the baseband complex envelope of the signal for compatibility with communication signals and to bring benefit of reduced sampling requirements.
- The model should be easy to train to allow fast adaptation. In particular, the feature of linear-in-parameter, meaning the output of the model is linear with respect to the modelling coefficients, is highly desirable as linear system identification algorithms such as least squares can be used to extract such coefficients.

Among the many models of nonlinear systems, there are only a few that display all of the above identified desirable features [67]. The Volterra series discussed previously fits perfectly as the ideal model for predistortion and has seen widespread usage. On the other hand, models based on ANN [45, 46, 47], Hammerstein [68] or Wiener structures [39], though powerful in terms of modelling capability, require complicated and costly training algorithms to obtain the coefficients.

As discussed before, a major disadvantage of the Volterra series formulation is that the number of kernels increases exponentially with nonlinearity order and memory depth. Thus, for ease of implementation as well as to reduce the complexity, many forms of polynomials-based predistortion models have been researched. These range from simple

memoryless models to more sophisticated schemes capable of handling significant memory effects [69, 70]. Memory polynomial (MP) [40] is a popular model, the baseband equivalent discrete representation being expressed as

$$\tilde{u}_{MP}[n] = \sum_{i=0}^M \sum_{j=0}^N a_{i,j} \tilde{x}[n-i] |\tilde{x}[n-i]|^j, \quad (2.14)$$

where $\tilde{x}[n]$ and $\tilde{u}[n]$ are baseband samples before and after the predistortion, $a_{i,j}$ are the complex model coefficients, N is the nonlinearity order, and M is the memory depth. It can be viewed as including only basis on the main diagonal axis of the Volterra series, and consists of multiple polynomials of the current and delayed signals.

The lack of cross terms from the delayed inputs in the MP formulation limits its performance and leads to over-modelling and numerical stability issues in practice. To address such issues, the generalized memory polynomial (GMP) formulation [41, 71] expands on the concept of MP to include more basis along the main diagonal, and introduces cross-terms as products of the current and lagging or leading envelopes, using both positive and negative time shifts. The general form of the GMP is expressed as

$$\begin{aligned} \tilde{u}_{GMP}[n] = & \sum_{k=0}^{K_a-1} \sum_{l=0}^{L_a-1} a_{kl} \tilde{x}[n-l] |\tilde{x}[n-l]|^k \\ & + \sum_{k=1}^{K_b} \sum_{l=0}^{L_b-1} \sum_{m=1}^{M_b} b_{klm} \tilde{x}[n-l] |\tilde{x}[n-l-m]|^k \\ & + \sum_{k=1}^{K_c} \sum_{l=0}^{L_c-1} \sum_{m=1}^{M_c} c_{klm} \tilde{x}[n-l] |\tilde{x}[n-l+m]|^k. \end{aligned} \quad (2.15)$$

Here, $K_a L_a$ are the number of coefficients for the aligned signal and envelope (memory polynomial); $K_b L_b M_b$ are the number of coefficients for the signal and lagging envelope; and $K_c L_c M_c$ are the number of coefficients for the signal and leading envelope. The GMP formulation achieves better performance at the cost of significant complexity and computational demands.

To improve linearization performance while maintaining relatively low complexity, the

complexity reduced Volterra (CRV) [43, 72] is an alternative scheme, formulated as follows

$$\begin{aligned}
\tilde{u}_n = & \sum_{\substack{i=0 \\ \text{even}}}^{N-1} a_{1,i,0} \tilde{x}_n |\tilde{x}_n|^i \\
& + \sum_{m=1}^M \sum_{\substack{i=0 \\ \text{even}}}^{N-3} \sum_{\substack{k=2 \\ \text{even}}}^{N-1} a_{m,i,k} \tilde{x}_n |\tilde{x}_n|^i |\tilde{x}_{n-m}|^k \\
& + \sum_{m=1}^M \sum_{\substack{i=0 \\ \text{even}}}^{N-1} \sum_{\substack{k=2 \\ \text{even}}}^{N-1} b_{m,i,k} \tilde{x}_{n-m} |\tilde{x}_n|^i |\tilde{x}_{n-m}|^k.
\end{aligned} \tag{2.16}$$

Pruning of the Volterra series can also be performed, as in dynamic deviation reduction Volterra (DDR-Volterra) [42]. It is based on eliminating Volterra bases with high dynamic order.

On the other hand, to simplify the implementation of the predistortion engine, the envelope memory polynomial (EMP) [73] is proposed as

$$\tilde{x}_{PD}[n] = a_{0,0} \tilde{x}[n] + \tilde{x}[n] \sum_{i=0}^M \sum_{\substack{j=2 \\ \text{even}}}^N a_{i,j} |\tilde{x}[n-i]|^j. \tag{2.17}$$

Its formulation is based on the envelope of the past signal only, and results in a much friendlier form for circuit implementation.

The introduction of MIMO in 5G with large numbers of PAs and antennas in a closely packed form creates new distortion mechanisms and has inspired new research efforts. Research work in [74, 75, 76, 77] focused on investigating the distortion mechanisms and forming new engine formulations to linearize each PA. The cross-over DPD model proposed in [74] models the crosstalk before the PA and directly cancels it and a different pruned memory polynomial is proposed for more efficient realization [75]. A dual-input with crosstalk and mismatch model (DI-CTMM) DPD is proposed by [77] to simplify the DPD formulation by using the CTMM to model the crosstalk. More recent studies [78, 79] have focused on linearizing the entire array instead of individual PA elements. They have demonstrated that, in the case of RF/hybrid beamforming arrays, a SISO DPD model can be used to linearize the main-beam signal using a far-field probing receiving antenna, making DPD a viable solution for mm-wave systems.

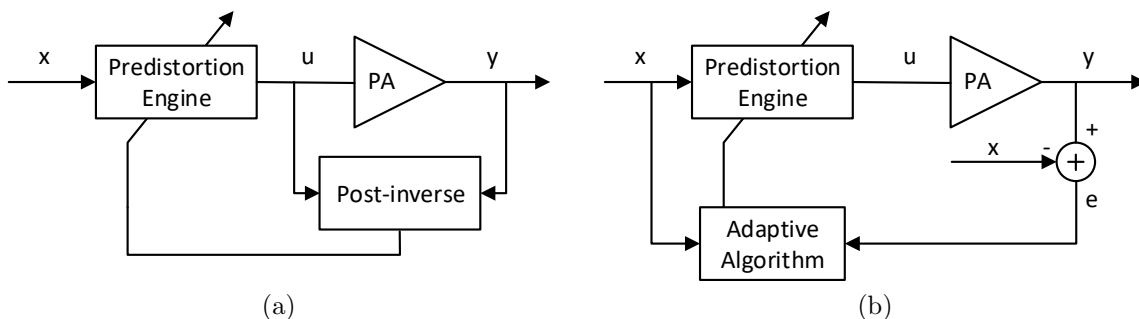


Figure 2.18: Block diagram of the indirect learning (a) and direct learning (b) architecture.

2.5.4 Feedback and Training

The feedback path and training algorithm are closely coupled to provide updates to the predistortion engine coefficients. Increasing bandwidth in communication signals has pushed the TOR towards its limits, as the ADCs struggles to provide sufficient sampling rate to capture the full bandwidth of the expanded PA distortion. Thus, new techniques to reduce the bandwidth and sampling rate requirements are under active investigation.

The training algorithm can be broadly categorized into direct and indirect learning architectures, as illustrated in Fig. 2.18. The indirect learning architecture, as shown in Fig. 2.18a, is based on the theoretical framework of p th-order inverses that shows the pre-inverse model is the same as the post-inverse for a nonlinear system [53]. The post-inverse model of the PA is first estimated using the PA's input and output. The coefficients for the post-inverse model are then used in the predistortion engine that implements the same model as the pre-inverse. An indirect learning algorithm can achieve good linearization performance after the first iteration, but iterations can be performed to converge on the optimal linearization results.

The direct learning architecture, as shown in Fig. 2.18b, tries to minimize the difference between the original input and the normalized output [80]. The output and the input are compared and the error is used to estimate the updates of the coefficients. Iterations are then performed until the coefficients converge to the optimal value that minimizes the measured error. A typical LSs estimator, commonly found in adaptive signal processing, such as the least mean square (LMS) or recursive least square (RLS) could be used.

Direct learning has seen increasing popularity in recent years. This is partly due to its better performance for wideband signals, as indirect learning is affected by the large amount of noise in PA output measurements that can lead to production of biased results.

A more attractive feature of the direct learning approach is its ability to enable sub-Nyquist sampling—allowing for a significant reduction in the ADC sampling rate in the TOR. In [81], the distortion of the feedback path is first characterized with the PA bypassed, and then used to restore the PA output from the under-sampled signal. This approach not only requires a PA-bypass circuit, but forward modelling of the PA is needed to characterize and compensate for the distortion of the feedback path. In [82, 83, 84], the PA output is filtered first to limit the bandwidth, reducing the ADC sampling rate required to capture it. This also reduces the linearization bandwidth of the DPD system, as part of the PA distortion is filtered out and not captured by the TOR. An alternative approach is proposed in [85, 86, 87, 88], where the ADC captures the aliased output with a sub-Nyquist sampling rate. Training algorithms are then used to extract the coefficients for the DPD engine from the sub-Nyquist signals and the full-rate input signal. Other works approach the problem by using only the real part of the signal to train the coefficients as presented in [89, 90].

MIMO systems present unique challenges to the feedback path, as there are many PAs and antennas in the system with PA crosstalk and antenna cross-coupling creating new mechanisms of distortion. In [77], the output of each PA is coupled out and captured separately, then used to train the CTMM that is fed to every dual-input DPD, resulting in very high system complexity. A similar TOR structure is used in [78], but the results are combined using digital anti-beamforming to compute the main-beam signal for training purposes. The other approach considers using an over-the-air (OTA) receiver to capture the far-field main beam of the array [79]. In [91], the same OTA setup is used to compute the integrated linear crosstalk model (ICTM), which is later used to compute the linear crosstalk signals for each predistortion block for each sub-array in hybrid beamforming. Using a receiver in the same array as a feedback is proposed in [92], with multiple near field antennas used as diversity feedback to estimate the combined output.

2.5.5 Hardware Implementation of Predistortion System

Besides theoretical works on predistortion function, feedback, and training methods, the hardware implementation of a real-time adaptive predistortion system also deserves attention. While research into ARFPD systems [59, 58, 60, 61, 65, 66] discussed previously inherently has closely coupled theory and implementation, in the more popular DPD systems the implementation details of the predistortion engine are often taken for granted and simulated in software. Except in the early dates in DPD development, advances in digital circuitry offer more than sufficient computational power for predistortion engines, and their hardware implementation has received limited research attention since. Yet, with the dramatically higher demands of 5G systems, particularly mm-wave systems, it becomes

difficult to satisfy the computation needs for DPD engines even with powerful high-speed DSPs built with advanced nano-scale process technology. To compensate for nonlinear distortion in RF PAs, the predistorted signal needs to be oversampled by a significant factor (typically 5 times the signal bandwidth) to compensate for the expanded bandwidth caused by intermodulation. For example, to support 5G NR mm-wave communication signals with >400 MHz modulation bandwidths, conventional DPD implementations would need digital circuits and data converters supporting bandwidths in the gigahertz, which are close to or beyond the limits of today’s digital circuits. Such requirements render the hardware implementation of real-time massive MIMO DPD difficult, and ultimately can result in major energy and cost overhead in the underlying hardware.

Several works have studied the hardware implementation of DPD systems with the goal of improving hardware efficiency, [93, 94, 95, 96, 97] being recent examples. In [93], a new technique for implementing Volterra series-based DPDs was proposed. This technique combines EMP and multiplication, together with time division multiplexing to reduce the hardware resources required. Alternatively, the authors in [94] proposed a DPD system with a reduced sampling rate using a new aliasing cancellation scheme for minimizing its power consumption at the expense of a slight increase in the complexity. Such work allows wideband signals to be linearized on an FPGAs with a lower clock frequency by focusing on the limited linearization bandwidth while sacrifice performance outside the linearization bandwidth. In [95], a DPD implementation using a feedback complexity-reduced Volterra-series (fCRV) model is described. By using a memoryless nonlinear block with an auxiliary memory compensation feedback filter, the overall hardware resource requirement can be significantly reduced. In [96], a multiple EMP DPD architecture based on the partial least-squares regression method is presented. It uses bilinear interpolation and extrapolation to implement the 3-D distributed memory EMP architecture. Finally, in [97], a cross-term enabled DPD model is implemented using piece-wise magnitude-selective affine functions as the nonlinear block, with significantly reduced resource usage. However, all of these works assume a serial digital input data stream for the DPD engine, effectively limiting their data rate to the maximum clock rate of the digital circuits. Modern system-on-chips (SOCs) based on FPGAs typically have data path maximum clock frequencies in the range of several hundred megahertz—far below the required rate for wideband DPD systems.

2.6 Summary

In this chapter, the use of linearization systems as a way to improve the efficiency of RF PAs were reviewed. The efficiency and linearity of PAs have been well studied. It has

been shown that conventional PAs will have low efficiencies when transmitting modern communication signals with high PAPR. Backoff efficiency enhancement techniques such as drain modulation and load modulation, as well as other efficiency enhancement methods like outphasing and digital PAs offer some benefits but further complicate the PA linearity problem as they introduce additional distortion mechanisms. At the same time, advancements in modern wireless communication networks are calling for RF PAs with high efficiency, ultra-wide bandwidth and very good linearity. Hence, PA linearization systems have been widely adopted as they compensate for various PA distortions and allow the PA to operate in the high efficiency region and use advanced architecture.

Various methods of PA linearization have been developed, among them the DPD system has seen wide adoption due to its excellent linearization capability, ease of implementation and ability to be integrated with the base system. Yet, advancements in wireless technology seeking small-cells with reduced PA output and the dramatically increased signal bandwidth in 5G mm-wave systems, pose new challenges to the DPD solution. In particular, the expanded bandwidth of the PA distortion and the corresponding predistortion function require a high oversampling rate in the DPD engine and TOR, incurring high power overhead and cost. A field of active research interest, the challenges are being tackled from multiple directions. There is renewed interest in analog/RF predistortion as a replacement for small-cell base stations, taking advantage of the low power nature of analog circuits. Yet system level designs that reduce hardware complexity, particularly for analog implementation, while maintaining good linearization capability are still needed. New formulations of the predistortion function continue to be investigated, with newer models being capable of handling complex distortion caused by wide signal bandwidths and interactions in MIMO systems. There are also significant developments in feedback and training methods, largely focused on breaking away from conventional Nyquist rate sampling of the PA output and enabling proper coefficient updating with significantly reduced ADC sampling rates in the TOR. Finally, the hardware implementation of the predistortion system has once again become an important research focus as the required computation of the predistortion engine has started to exceed the limits provided by the digital circuitry available.

Chapter 3

Digitally Assisted Analog Radio Frequency Predistortion System

3.1 Introduction

In conventional base stations, RF PAs usually use advanced efficiency enhancement techniques and operate in the high efficiency low linearity region. Distortions within the PA are usually corrected for using DPD systems, for which the power consumption remains constant regardless of PA output power. This is a desirable feature for macro-cell base stations with PA output powers of hundreds of Watts. However, the push for higher data rates has led to shrinking cell sizes and small-cell base stations with reduced output powers in the range of a few Watts. The constant power overhead of DPD systems becomes a disadvantage instead in this case, reducing the overall efficiency of the transmitter system.

To address the challenges imposed by small-cell base stations in modern communication networks, and to overcome the limitations of existing PA linearization solutions, new architectures for low-power wideband PA linearization are required. Analog and RF predistortion techniques have received renewed attention thanks to their low power consumption. Yet, conventional analog or RF predistortion techniques [58, 59, 60, 61] have limited linearization performance and lack a viable solution to address memory effects. The ARFPD technique [65, 66] shows promising results through the use of an analog predistortion engine, but its linearization performance is still hindered by the difficulty of implementing a predistortion model of comparable linearization capacity to DPD using analog hardware.

This chapter attempts to address such problems by introducing the DA-ARFPD architecture with a novel predistortion model that allows for a reduction of hardware implemen-

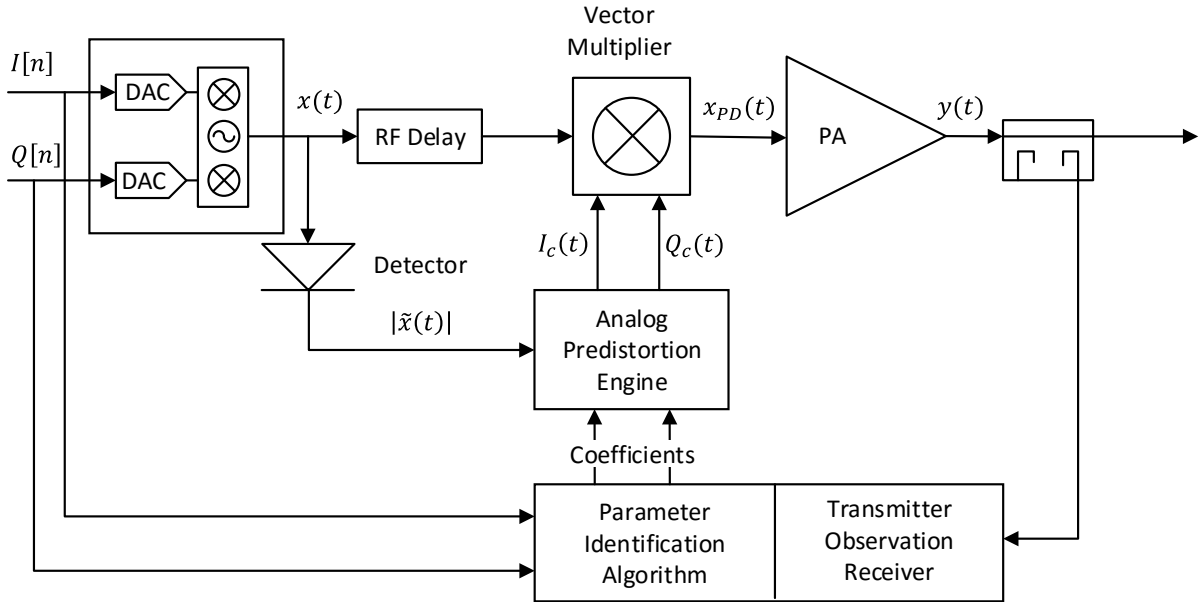


Figure 3.1: Block diagram of the analog/RF predistortion system.

tation complexity for the analog predistortion engine while maintaining good linearization capacity and low power overhead for sub-6 GHz small-cell base stations. While the proposed model is no longer linear-in-parameter in general, a new parameter identification algorithm is proposed for the training module to allow linear time computation of the model coefficients. A DA-ARFPD test bench, including the key components such as IQ modulators and RF vector multiplier, is devised to realistically assess the performance of the proposed model and training algorithm in the presence of actual analog hardware.

3.2 Linear Filter Assisted Envelope Memory Polynomial Model Overview

In an ARFPD system, as shown in Fig. 3.1, the generation of predistorted signals usually involves two steps: a) synthesis of a nonlinear function (e.g., MP) that is complementary to the distortion of the PA and b) application of the synthesized predistortion function output to the RF input signal to generate the required predistorted outputs.

One popular approach to address the memory effects is to use the MP model whose conventional baseband equivalent discrete representation is expressed in (2.14) Reformu-

lating the MP of (2.14) in the continuous time domain for the ARFPD scheme, the RF predistortion signal $u(t)$ can be expressed as

$$\begin{aligned} u(t) &= \text{Re}[\tilde{u}(t)e^{j\omega t}] \\ &= \sum_{i=0}^M \text{Re}[\tilde{x}(t - \tau_i) f_i(|\tilde{x}(t - \tau_i)|) e^{j\omega t}] \\ &= \sum_{i=0}^M \text{Re}[\tilde{x}(t - \tau_i) e^{j\omega t} f_i(|\tilde{x}(t - \tau_i)|)] \end{aligned} \quad (3.1)$$

$$\text{with } f_i(|\tilde{x}(t - \tau_i)|) = \sum_{j=0}^N a_{i,j} |\tilde{x}(t - \tau_i)|^j, \quad (3.2)$$

where $f_i(\cdot)$ is a nonlinear function corresponding to a memory term of order i , and τ_i denotes the time delay. The block diagram corresponding to (3.1) is illustrated in Fig. 3.2. The envelope signals $|\tilde{x}(t)|$ are generated by the envelope detector followed by analog delay elements (D), and the delayed RF inputs are generated using RF delay elements (RF-D). The many RF-D elements and vector multipliers required for this architecture would consume a large amount of physical space and increase the hardware complexity. To simplify the hardware implementation, attempts have been made to remove the dependency of the memory paths on receiving phase information from the preceding input signals by using the envelopes of the signals only. The EMP in [73] is one such approach. The baseband complex output $\tilde{u}[n]$ is given by (2.17). A reformulation of (2.17) for ARFPD results in:

$$u(t) = \sum_{i=0}^M \text{Re}[\tilde{x}(t) f_i(|\tilde{x}(t - \tau_i)|) e^{j\omega t}]. \quad (3.3)$$

The performance of the EMP predistorter is assessed using a wideband GaN DPA. Measurement results indicate that the performance of the EMP model is adequate when the signal bandwidth is limited to 20 MHz, but it significantly degrades as the modulation bandwidth extends beyond that. Such shortcomings of EMP model are attributed to its limited ability to model the linear memory distortion. To illustrate the point, the measured AM/AM and AM/PM plots of the GaN DPA, and those predicted by EMP, are shown in Fig. 3.3[98]. It is clear that the EMP model results have good agreement with the measured data at high input powers where the nonlinearity of the PA is more pronounced, but fails to predict the distortion at low powers where the linear memory distortion dominates (also where there is wider signal bandwidth).

This limitation of EMP is a result of the simplification used to derive its expression from the Volterra series. The simplification assumes flat behaviour by the fundamental

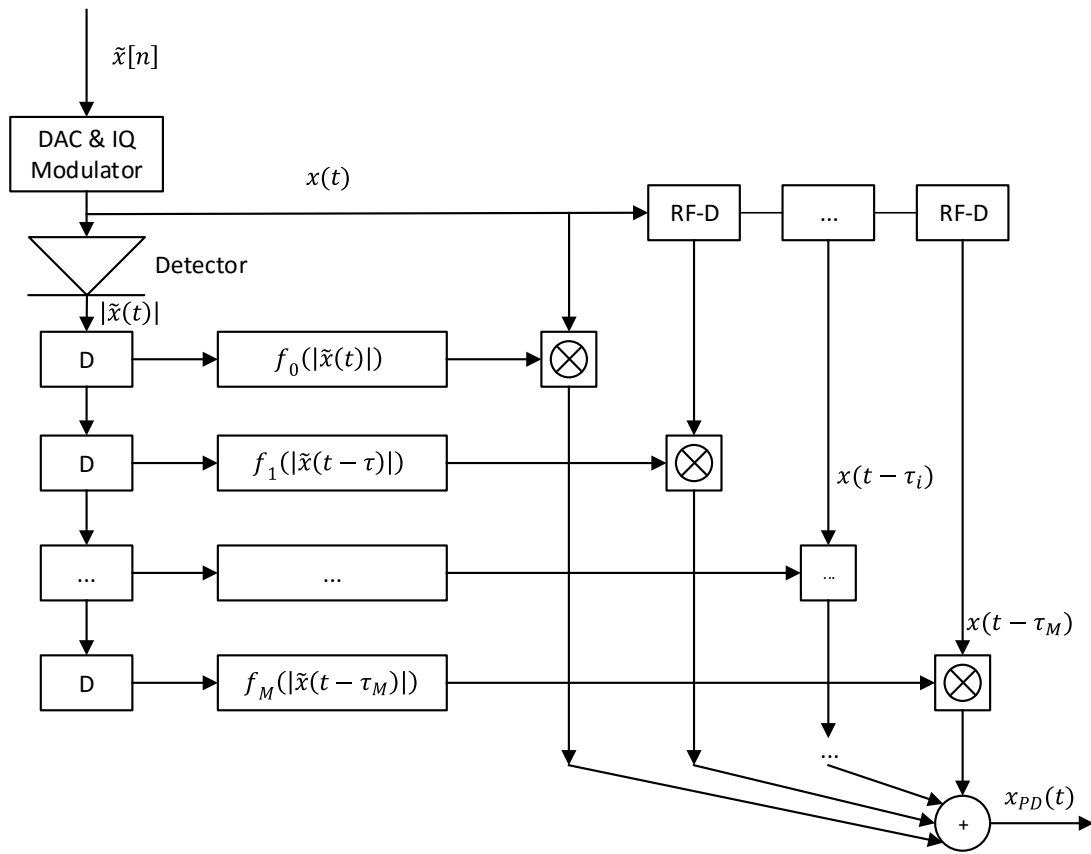


Figure 3.2: Block diagram of the MP function when applied to the ARFPD.

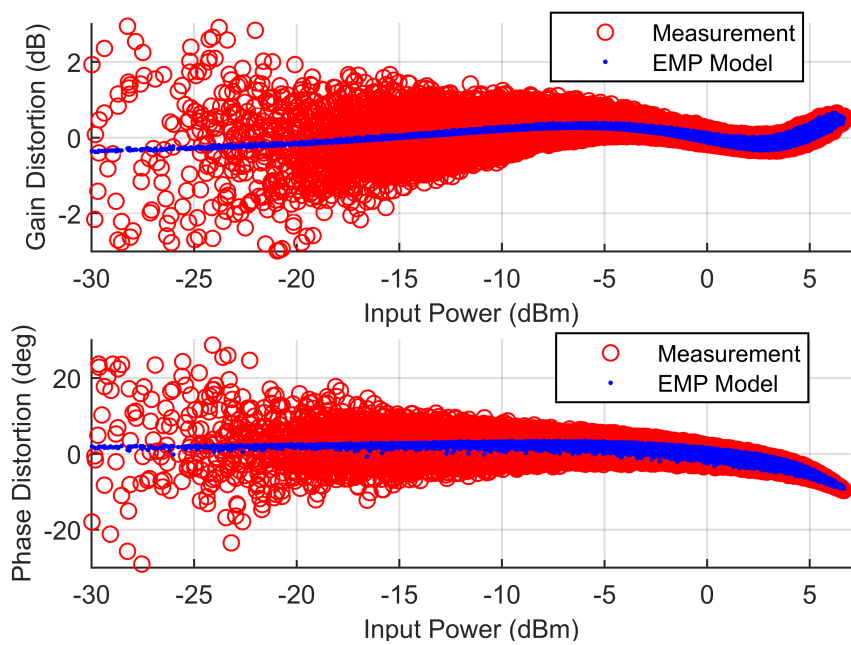


Figure 3.3: AM/AM and AM/PM characteristics, measured and modelled using the EMP model of a GaN DPA driven with a 20 MHz WCDMA signal at 2 GHz.

and its harmonics [99], and that the passband of the PA is much larger than the RF signal bandwidth [100]. Wider modulation bandwidth signals challenge this narrowband assumption. To accurately model and compensate for the memory effects, both the amplitude and phase information of preceding signals are required[100].

To address the shortcomings of EMP, linear filter is added before the EMP engine [98]. As shown in the block diagram in Fig. 3.4, the linear filter is implemented as a finite impulse response (FIR) filter in digital baseband. The resultant signal from the linear filter is then predistorted using EMP to compensate for the static and dynamic nonlinearities.

The advantages of the proposed FIR-EMP model in the DA-ARFPD setup over the conventional MP model can be seen clearly by comparing Fig. 3.2 and Fig. 3.4. The FIR filter in the digital baseband can be clocked at a much lower speed than a full DPD, thus requiring lower power overhead, as it must only cover the bandwidth of the baseband signal rather than the typical five times factor imposed by a DPD. EMP is also much simpler to realize in analog hardware than the MP based predistorter as it does not require multiple RF delay elements and vector multipliers.

3.3 FIR-EMP Parameter Identification Algorithm

Challenges in Identifying the Parameters

In the proposed FIR-EMP scheme outlined in the previous section, the FIR filter block is implemented in the digital domain and the EMP block is implemented in the RF domain. The cascade of a linear filter and a highly nonlinear block makes identifying the respective coefficients a major challenge.

The commonly used LS algorithm, popular for identifying the coefficients of a single-block predistortion model (e.g., MP or low pass equivalent Volterra), cannot be used to identify the respective coefficients of such a two-block nonlinear system due to the lack of linear relationships between the coefficients and the output signals. One potential solution is to use a nonlinear optimization algorithm such as the quasi-Newton method but this can be computationally intensive and is not suitable for real-time applications.

Proposed Identification Algorithm

In this section, a small-signal-assisted parameter identification (SSAPI) algorithm is proposed to efficiently identify the coefficients of the FIR filter and the EMP block, respectively

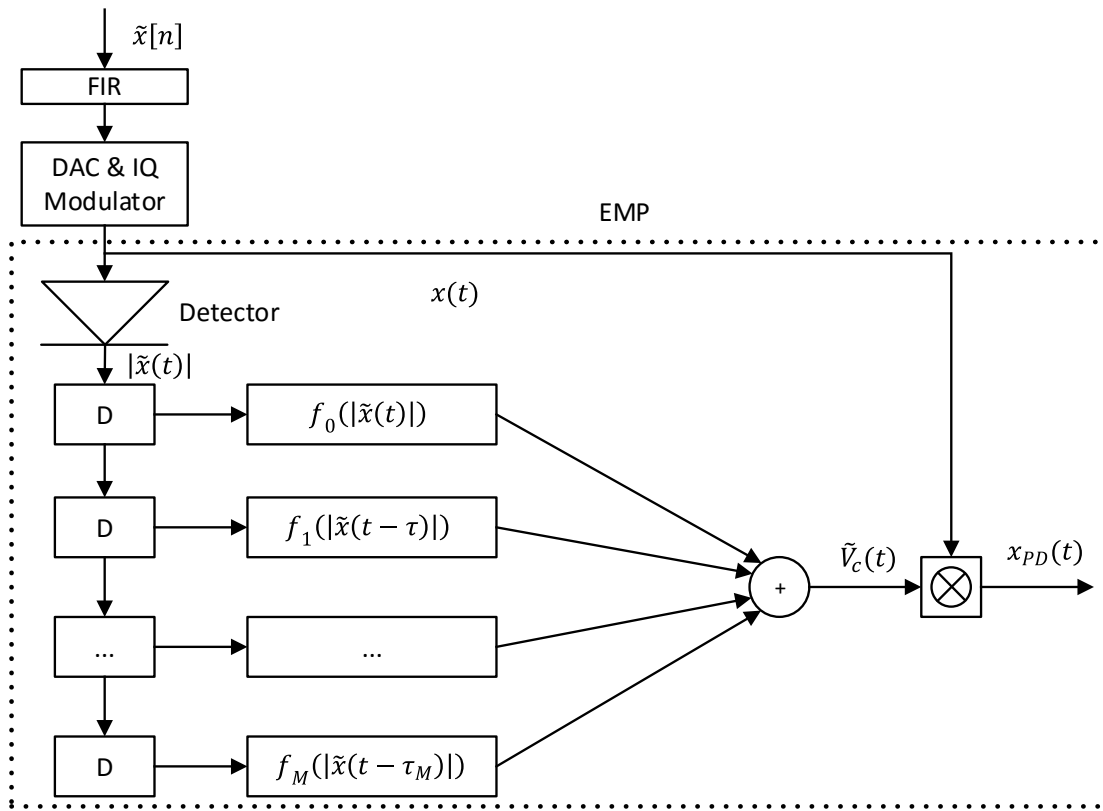


Figure 3.4: Block diagram of the proposed FIR filter assisted EMP function when applied to the DA-ARFPD.

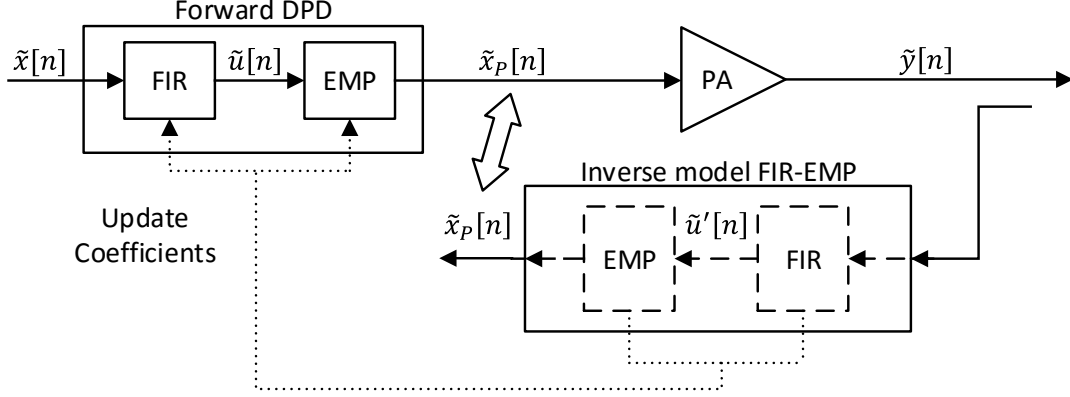


Figure 3.5: Training scheme for identifying the coefficients of the FIR-EMP predistorter.

without resorting to nonlinear optimization.

The sample envelope of the desired predistortion signal, $\tilde{x}_p[n]$, is generated by the forward DPD using the two block FIR-EMP as

$$\tilde{x}_p[n] = \tilde{u}[n] \sum_{i=0}^M \sum_{j=0}^N a_{ij} |\tilde{u}[n-i]|^j$$

$$\text{with } \tilde{u}[n] = \sum_{k=0}^K c_k \tilde{x}[n-k]. \quad (3.4)$$

The forward DPD can be assumed to be the same as the post-distortion FIR-EMP block, which is the inverse model of the PA, taking samples of the PA output envelope $\tilde{y}[n]$ and outputting $\tilde{x}_p[n]$, as depicts in Fig. 3.5. Hence, the coefficients of the EMP a_{ij} and FIR filter c_k in (3.4) can be found using the following relationship

$$\tilde{x}_p[n] = \tilde{u}'[n] \sum_{i=0}^M \sum_{j=0}^N a_{ij} |\tilde{u}'[n-i]|^j$$

$$\text{with } \tilde{u}'[n] = \sum_{k=0}^K c_k \tilde{y}[n-k], \quad (3.5)$$

where $\tilde{u}'[n]$ is the intermediate training data at the output of the FIR filter block, K is the

order of the FIR filter. Expanding (3.5) leads to the following expression

$$\tilde{x}_p[n] = \left(\sum_{k=0}^K c_k \tilde{y}[n-k] \right) \sum_{i=0}^M \sum_{j=0}^N a_{ij} \left| \sum_{k=0}^K c_k \tilde{y}[n-k-i] \right|^j. \quad (3.6)$$

According to (3.6), estimating the values of c_k and a_{ij} generally requires an advanced nonlinear optimization algorithm, which is possible but impractical in terms of the required time and computation resources.

In the proposed SSAPI algorithm, the basic principle is to use a small-signal $\tilde{x}_{pss}[n]$ to probe the linear memory effects of the nonlinear device, which avoids the static nonlinearity and the nonlinear memory effects associated with the device, and leaves the linear memory effects the dominant source of distortion. To illustrate this, assume that the magnitude of the output $\tilde{y}_{ss}[n]$ is small enough such that all of the higher order terms (e.g., $j \geq 1$) can be approximated as zero. Equation (3.6) can be simplified to

$$\begin{aligned} \tilde{x}_{pss}[n] &\cong \left(\sum_{k=0}^K c_k \tilde{y}[n-k] \right) \sum_{i=0}^M \sum_{j=0}^N a_{ij} \left| \sum_{k=0}^K c_k \tilde{y}[n-k-i] \right|^j \\ &\cong \left(\sum_{k=0}^K c_k \tilde{y}[n-k] \right) \sum_{i=0}^M a_{i0} \\ &\cong a_0 \sum_{k=0}^K c_k \tilde{y}_{ss}[n-k], \end{aligned} \quad (3.7)$$

where $a_0 = \sum_{i=0}^M a_{i0}$ is the linear gain, and can be absorbed by the FIR coefficients c_k . With (3.7), coefficients of the FIR filter block are estimated using the LS algorithm.

In order to obtain the small-signal training data, $\tilde{x}_{pss}[n]$ and $\tilde{y}_{ss}[n]$, one *obvious* approach is to operate the PA at a significant backoff region, where the linear memory effects dominate and the nonlinearity of the PA can generally be neglected. However, this *obvious* approach has three limitations. First, the linear memory effects of the PA might change due to a shift in the PA's region of operation. Second, the small-signal stimulus requires a high dynamic range in the transmitter observation receiver and the recorded samples could be sensitive to measurement noise. Lastly, forcing the PA to operate in the small-signal region necessitates off-line training and is not suitable for field applications.

In consideration of the aforementioned limitations, this *obvious* approach is deemed unsuitable for obtaining the small-signal training data. To address such limitations, it is

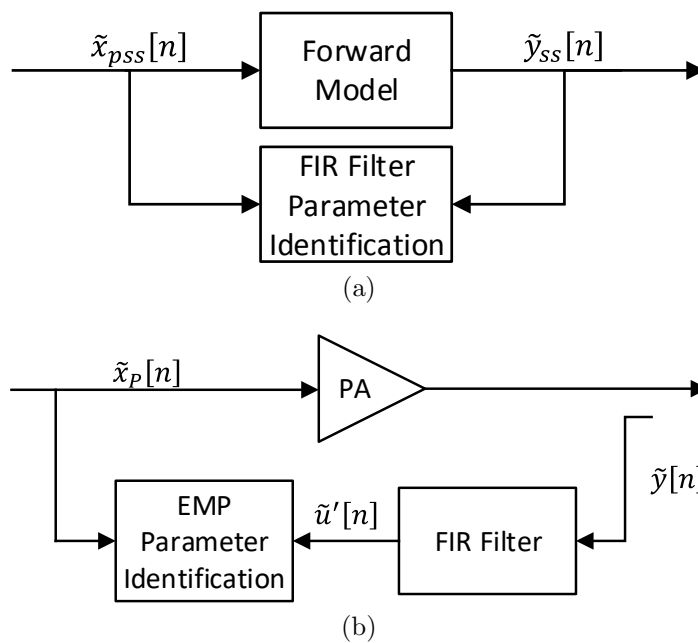


Figure 3.6: Steps of the proposed SSAPI algorithm for the FIR-EMP predistorter (a) FIR parameter identification using the forward model of the PA (b) EMP parameter identification using the intermediate output of the FIR block.

proposed that a forward model (e.g., Volterra) of the PA is estimated first, and the small-signal training data (e.g., $\tilde{x}_{pss}[n]$ and $\tilde{y}_{pss}[n]$) can be deduced using the forward model. As the small-signal training data are obtained using the forward model in the software without forcing the PA to operate at a significant backoff region, the aforementioned limitations are less problematic. Fig. 3.6 summarizes the steps of the proposed SSAPI algorithm and the detailed algorithm is presented below:

1. Estimate the coefficients of the PA forward model, using the conventional LS algorithm from the sampled input and output of $\tilde{x}_p[n]$ and $\tilde{y}[n]$.
2. Compute the small-signal output $\tilde{y}_{ss}[n]$, from the forward model, using a small-signal stimulus $\tilde{x}_{pss}[n]$ as the input, as shown in 3.6(a).
3. Estimate the coefficients of the FIR block c_k , using the conventional LS algorithm based on $\tilde{x}_{pss}[n]$ and $\tilde{y}_{ss}[n]$.
4. Compute the intermediate output of the FIR block $\tilde{u}'[n]$, according to (3.5).
5. Estimate the coefficients of the EMP block a_{ij} , using the conventional LS algorithm based on $\tilde{u}'[n]$ and $\tilde{x}_p[n]$, as shown in 3.6(b).

Evaluation of the Proposed SSAPI Algorithm

In order to assess the validity of the proposed parameter identification algorithm, coefficients of the FIR-EMP predistorter are estimated using 1) the quasi-Newton algorithm from the MATLAB nonlinear optimization toolbox and 2) the proposed SSAPI algorithm. The training data is the same test data used in Fig. 3.3. Out of the 50000 available points, 10000 points are used to train the proposed FIR-EMP model and the other 40000 points are used for model validation purpose.

Fig. 3.7 compares the modelling NMSE versus the number of iterations corresponding to the two algorithms. Note that each of the two algorithms takes approximately the same computation time per iteration. According to Fig. 3.7, the proposed SSAPI algorithm achieved an NMSE of -41.8 dB and only requires one iteration. In contrast, the nonlinear optimization requires ~ 14000 iterations to reach the same level of modelling accuracy.

It is worth mentioning that the described SSAPI algorithm is specifically proposed for the FIR-EMP model and is fundamentally different from the well-known parameter identification algorithm used by the Hammerstein or Wiener models [101, 68]. The two-box Hammerstein or Wiener model that consists of a static nonlinearity and an FIR filter

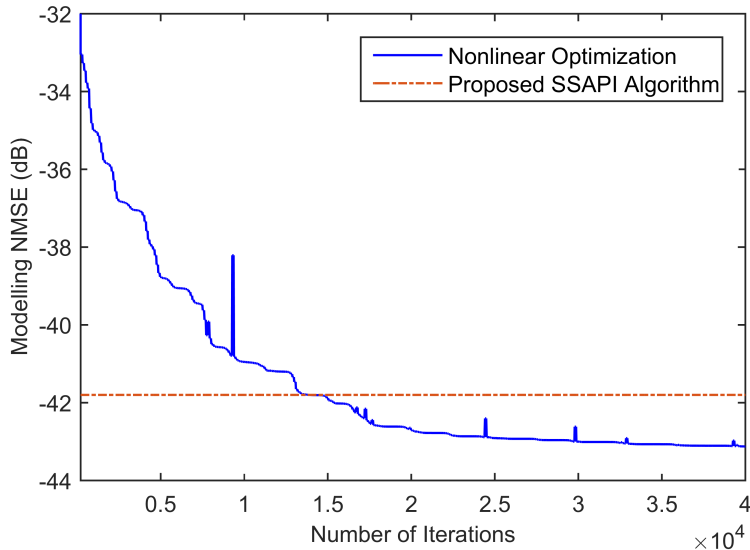


Figure 3.7: Modelling accuracy vs. the number of iterations for the quasi-Newton nonlinear optimization and the proposed SSAPI algorithm.

is typically trained by first identifying the static nonlinearity, either through a CW test [101] or a moving average method [68], before the FIR block can be identified. However, in the case of the proposed FIR-EMP model, the linear and nonlinear memory effects modelled by the FIR and EMP blocks respectively, are coupled and cannot be separated for conventional CW testing or the moving average method. On the other hand, the proposed SSAPI algorithm begins by first identifying a PA forward model, which is then used to generate new sets of input-output signals at the small-signal, so that the FIR filter coefficients can be identified. The coefficients of the EMP block are then identified using typical LS algorithms.

3.4 Validation and Measurement Results

To realistically assess the performance of the proposed FIR-EMP model in an DA-ARFPD system, a test bench including the key component is devised. Its corresponding block diagram is shown in Fig. 3.8 and the photograph of the test bench is shown in Fig. 3.9. The test bench consists of five main parts: the DSP unit, IQ modulator, RF vector multiplier, PA under test and the transmitter observation receiver.

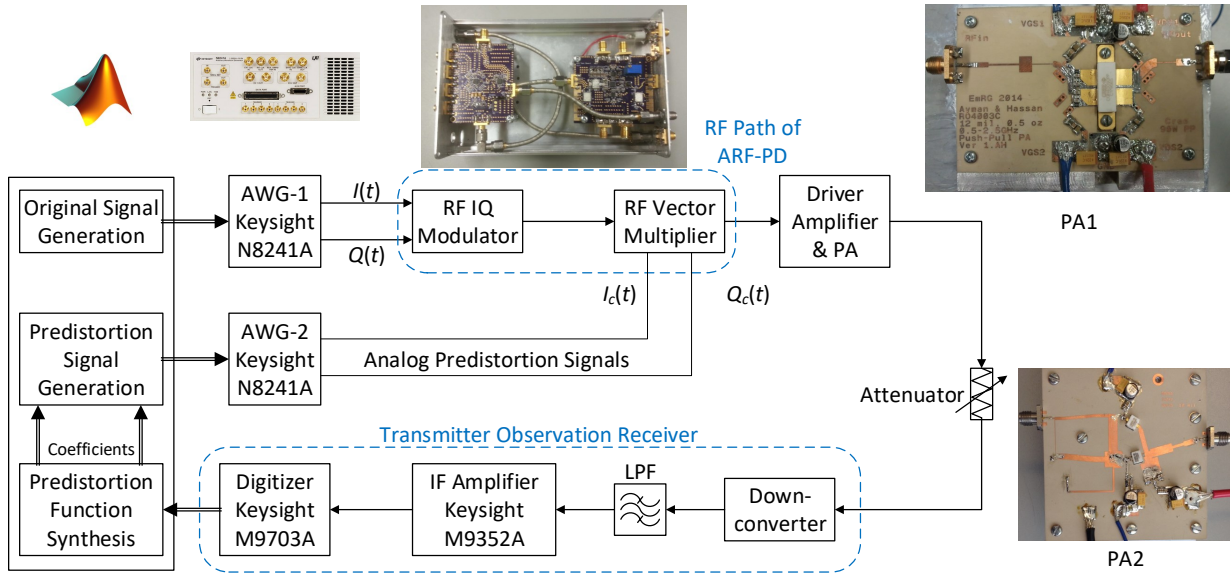


Figure 3.8: Block diagram of the DA-ARFPD engine test bench and the PAs under test.

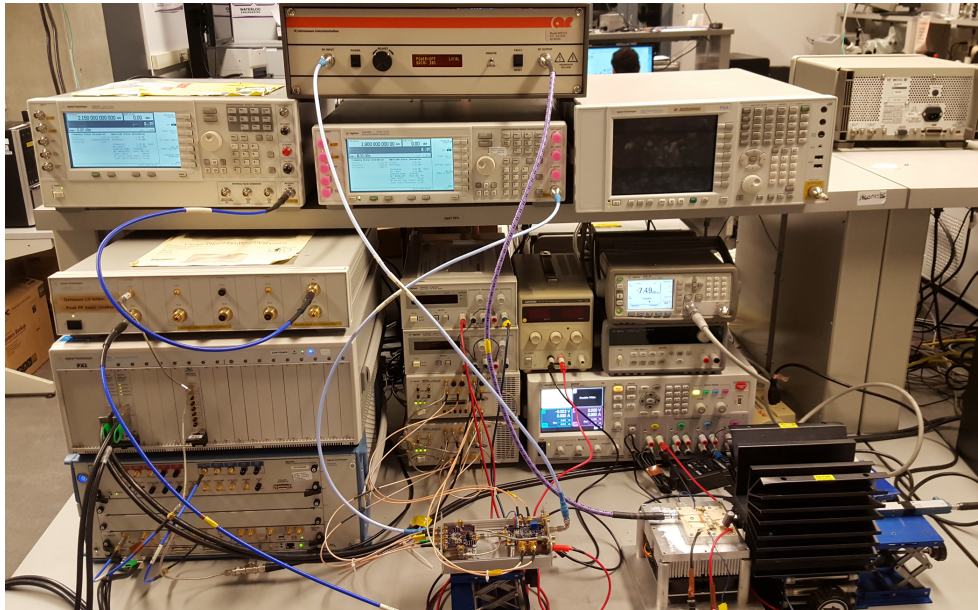


Figure 3.9: Photograph of the DA-ARFPD measurement setup.

The DSP unit synthesizes the predistortion function (i.e., FIR-EMP) and generates the corresponding signals in the digital baseband. Two Keysight N8241A arbitrary waveform generators (AWGs) are synchronized using the same sampling clock (625 MHz). AWG-1 sends the baseband analog in-phase and quadrature signals $I(t)$ and $Q(t)$ to the transmitter (ADL5375) for up-conversion. AWG-2 generates baseband signals $I_c(t)$ and $Q_c(t)$ and sends them to the RF vector multiplier (ADL5390) in order to apply the predistortion function to the incoming RF signal. The output of the PA under test is captured using a Keysight N9030A signal analyzer with a maximal observation bandwidth of 160 MHz. All the equipment is synchronized using a 10 MHz reference clock.

Compared to a complete DA-ARFPD system, the current test bench implements the major RF building blocks (i.e., IQ modulator and RF vector multiplier) but uses an emulated predistortion engine. Although implementation of the predistortion engine in hardware is feasible, this function would require a dedicated integrated circuit design. As a proof-of-concept evaluation, the described DA-ARFPD test bench is used to evaluate the proposed FIR-EMP model at the system level.

The proposed FIR-EMP model and the proposed SSAPI algorithm are evaluated in the DA-ARFPD test bench described previously to linearize different PAs driven by various wideband and intra-band carrier-aggregated signals. Two PAs under test are used, consisting of:

1. A GaN push-pull PA with 85-W peak envelope power operating at a carrier frequency of 900 MHz (PA1) [102]
2. A GaN DPA with 20-W peak envelope power operating at a carrier frequency of 1.9 GHz (PA2)[103].

The nonlinearity order is found out to be 7 and the memory depth is found out to be 5 based on the PAs characteristics through iterative parameter sweeping.

For comparison purposes, a DPD test bench is also built. This DPD test bench is identical to Fig. 3.8 but without the RF vector multiplier and AWG-2. Three predistortion test cases are designed:

1. using the MP model in the DPD setup
2. using the dynamic deviation reduction based Volterra series (DDR-Volterra) model [42] in the DPD setup
3. using the proposed FIR-EMP model in the DA-ARFPD setup.

Table 3.1: Summary of the Linearization Performance of PA1 Driven by Wideband Modulated Signals

PA Under Test		85 Watt Push-Pull GaN PA (900 MHz)		
Test Signals		20 MHz	30 MHz	40 MHz
Average Power (dBm)		38.2	38.8	38.6
Without PD	ACLR	-32.8/-34.2	-34.2/-33.8	-34.7/-34.7
	EVM	9.4%	8.0%	9.9%
DDR-Volterra (DPD test bench)	ACLR	-55.1/-56.3	-57.5/-57.5	-52.4/-54.0
	EVM	0.61%	0.52%	0.61%
MP (DPD test bench)	ACLR	-49.2/-49.6	-51.7/-51.8	-47.0/-47.1
	EVM	0.78%	0.71%	0.95%
FIR-EMP (DA-ARFPD test bench)	ACLR	-52.2/-51.8	-54.4/-54.2	-50.1/-51.3
	EVM	0.69%	0.67%	0.82%

The validity of the proposed FIR-EMP model is first assessed in the DA-ARFPD test-bench to linearize PA1 and PA2, driven by two different wideband modulated signals:

1. A 20 MHz LTE-A signal with a PAPR of 8.9 dB
2. A 30 MHz 4-carrier 110011 wideband code division multiple access (WCDMA) signal with a PAPR of 8.6 dB.

Table 3.1 and Table 3.2 summarize the linearization results for PA1 and PA2, respectively, corresponding to the different test cases. The PA distortions (AM/AM and AM/PM) corresponding to PA1 and PA2 driven under the 30 MHz signal are plotted in Fig. 3.10 and 3.11. The output spectra of PA1 and PA2 under the three test cases are plotted in Fig. 3.12 and 3.13.

For PA1, the proposed FIR-EMP model in the DA-ARFPD setup achieved an ACLR improvement of 20 dBc when linearizing a 20 MHz LTE-A signal. Compared with the MP model based DPD setup, up to 2.7 dB ACLR improvement is noticed.

Table 3.2: Summary of the Linearization Performance of PA2 Driven by Wideband Modulated Signals

PA Under Test		20 Watt Doherty GaN PA (1900 MHz)			
Test Signals		20 MHz	30 MHz	40 MHz	80 MHz
Average Power (dBm)		34.2	34.8	34.6	32.5
Without PD	ACLR	-32.9/-30.9	-32.1/-30.5	-34.6/-31.6	-34.8/30.8
	EVM	7.4%	7.7%	6.2%	9.6%
DDR-Volterra (DPD test bench)	ACLR	-52.4/-51.9	-54.1/-52.5	-51.2/-49.8	-49.8/-48.5
	EVM	0.87%	0.71%	0.82%	1.4%
MP (DPD test bench)	ACLR	-47.6/-48.7	-50.2/-50.8	-46.7/-48.1	-45.0/-43.8
	EVM	0.90%	1.20%	1.13%	3.5%
FIR-EMP (DA-ARFPD test bench)	ACLR	-53.1/-52.9	-54.8/-54.6	-50.3/-50.2	-48.0/-47.7
	EVM	0.76%	0.70%	0.85%	1.4%

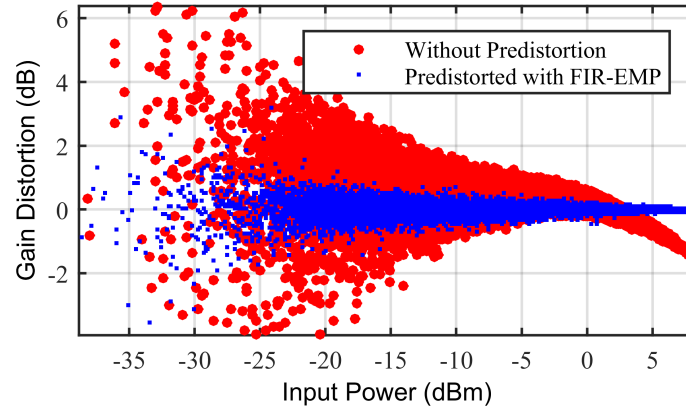
For PA2, the advantages of the proposed model are evident according to Table 3.2. The ACLR measured using the proposed FIR-EMP are 4.6 dB better than the MP model based DPD for the 30 MHz WCDMA signal. Linearization results from the DDR-Volterra model are provided as the baseline for comparison, and the FIR-EMP model in the DA-ARFPD test bench achieves comparable linearization capacity.

Fig. 3.13 shows the excellent linearization capacity of the proposed FIR-EMP model by comparing the spectra recorded for PA2. The proposed FIR-EMP test case has significantly lower in-band and out-of-band distortion compared to the spectra obtained using the MP based DPD. It achieves comparable linearization results to the sophisticated DDR-Volterra model based DPD setup.

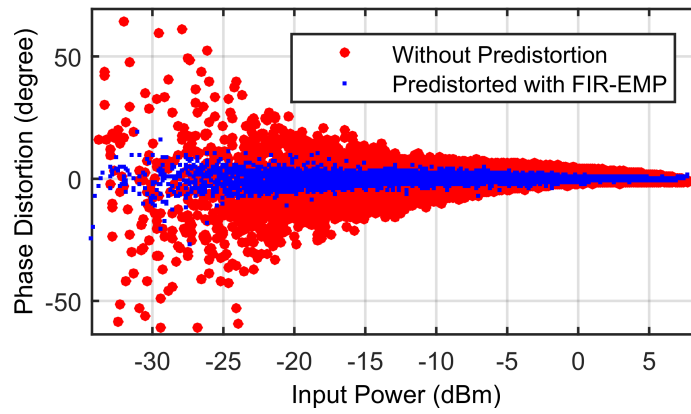
To further evaluate the performance of the proposed FIR-EMP model under newer 4G communication signals, a 40 MHz mixed standard carrier aggregated signal, consisting of a 15 MHz 3-carrier WCDMA signal and a 15 MHz LTE-A signal with a combined PAPR of 8.4 dB is synthesized.

For PA1, the proposed FIR-EMP in the DA-ARFPD test-bench achieves an ACLR of -50.1 dBc, compared to an ACLR of -47 dBc found using the MP in DPD. For PA2, the proposed FIR-EMP test case achieve an ACLR of -50.2 dBc, compared to the MP test case which records an ACLR of -46.7 dBc. Compared with the sophisticated DDR-Volterra model, the proposed FIR-EMP model achieves a similar level of linearization (i.e., about -50 dBc ACLR).

To further assess the performance of the proposed FIR-EMP under even wider bandwidths, an 80 MHz mixed-standard intra-band carrier aggregated signal, consisting of a 20 MHz 4-carrier WCDMA signal and a 20 MHz LTE-A signal with a combined PAPR of 9.5 dB. As summarized in Table 3.2, when compared with the MP test case, the proposed FIR-EMP model achieves a 4 dB improvement in the measured ACLR. Significant in-band distortion is observed in the MP based DPD case, which leads to a high EVM of 3.5% while the proposed FIR-EMP model has a measured EVM of 1.4%. The output spectra are presented in Fig. 3.14.

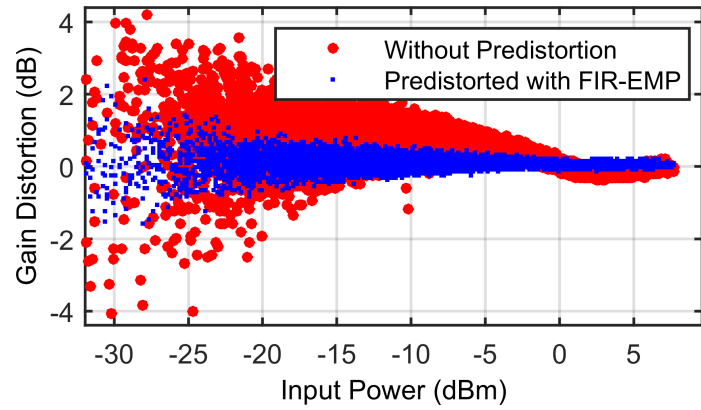


(a)

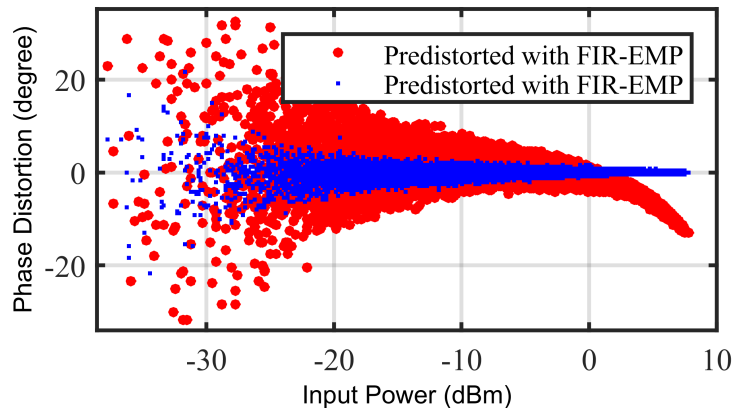


(b)

Figure 3.10: (a) AM/AM and (b) AM/PM characteristics for PA1 without predistortion (red) and with the proposed FIR-EMP in DA-ARFPD test bench (blue) of a 30 MHz WCDMA signal.



(a)



(b)

Figure 3.11: (a) AM/AM and (b) AM/PM characteristics for PA2 without predistortion (red) and with the proposed FIR-EMP in DA-ARFPD test bench (blue) of a 30 MHz WCDMA signal.

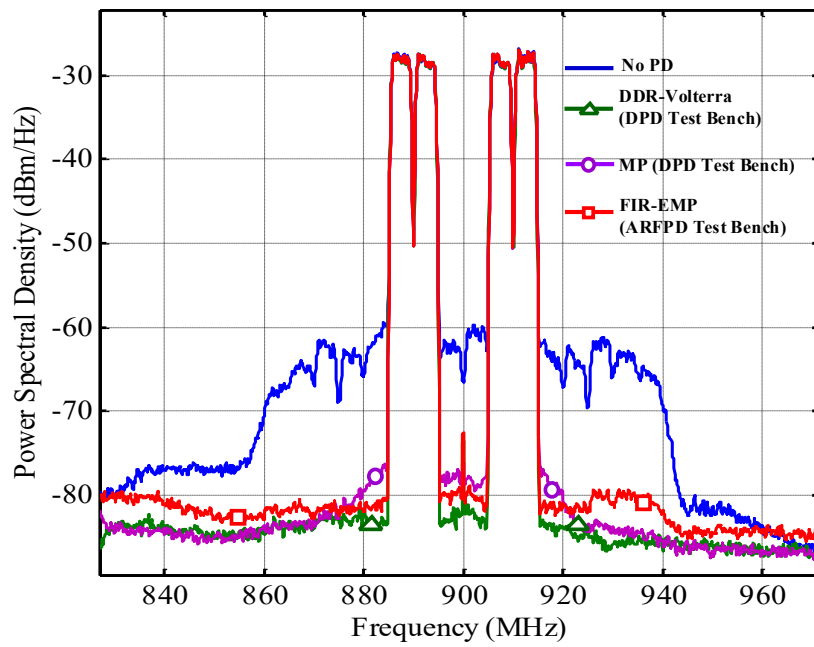


Figure 3.12: Output spectra of PA1 driven with 30 MHz bandwidth signal.

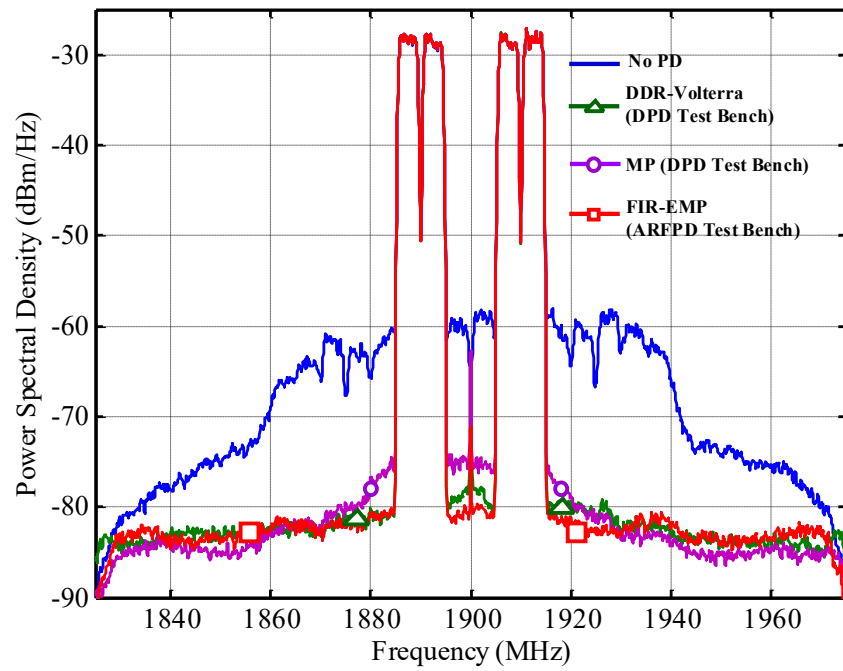


Figure 3.13: Output spectra of PA2 driven with 30 MHz bandwidth signal.

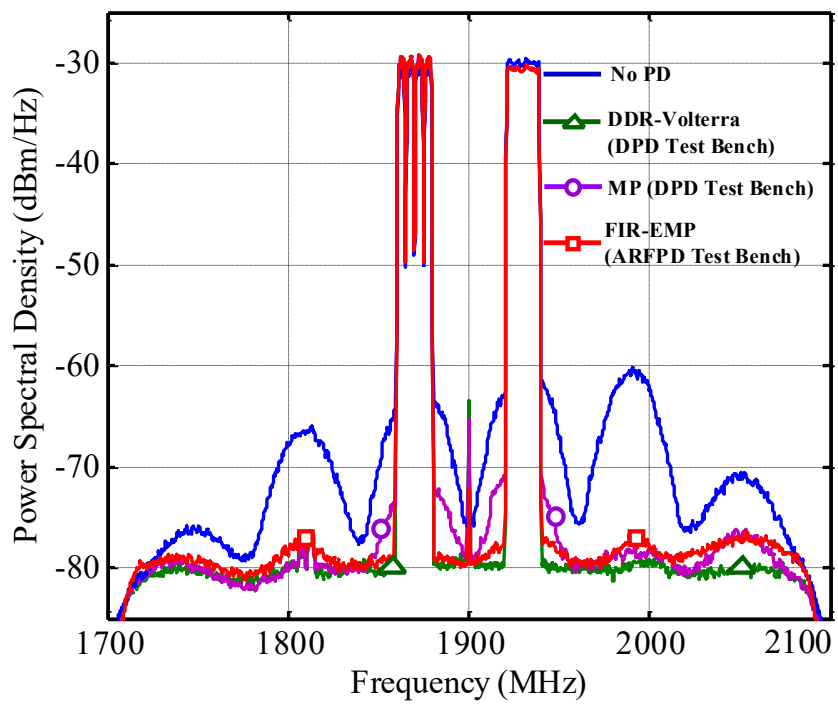


Figure 3.14: Output spectra of PA2 driven by 80 MHz signal.

3.5 Conclusion

In this chapter, a DA-ARFPD system using the FIR-EMP model along with a linear SSAPI algorithm were presented. A DA-ARFPD test bench, incorporating the major RF components, was built to assess the validity of the proposed FIR-EMP scheme and the SSAPI algorithm. It was demonstrated that the SSAPI algorithm can extract coefficients with excellent modelling accuracy for the cascaded blocks of the FIR-EMP in a single iteration. Measurement results have shown that the proposed FIR-EMP model using the SSAPI algorithm can successfully linearize multiple PAs driven with various wideband and CA signals of up to 80 MHz instantaneous bandwidths. Linearization performance comparable to a DDR-Volterra based DPD scheme indicates the viability of the proposed FIR-EMP model for implementing DA-ARFPD modules capable of mitigating the distortions exhibited by PAs driven by communication signals with up to 80 MHz modulation bandwidths. This confirms the potential of the DA-ARFPD approach as a very promising candidate for the linearization of sub-6 GHz small-cell base station PAs, reducing the power overhead compared to conventional DPD techniques.

Chapter 4

Real-time Digital Predistortion Hardware Architecture for Enabling Wideband Linearization of 5G Transmitters

4.1 Introduction

One of the primary aims of 5G communication networks is to support enhanced mobile broadband–high-speed wireless communication at data rates reaching over gigabits per second. To meet the demands of such high data rates, 5G networks are expected to use spectrum-efficient ultra-wideband signals with bandwidths in the range of hundreds of megahertz at mm-wave frequencies. Massive MIMO systems that integrate a large number of antennas, PAs and other RF elements will be the key technology to support efficient transmission of such signals. Yet, with large numbers of PAs and close form factors, conventional DPD solutions designed for single PAs are usually not practical. Studies in [78, 79] have shown that a SISO DPD model can be used to linearize a mm-wave MIMO transmitter, making DPD a viable solution for mm-wave systems in theory.

Yet, practical implementation of such DPD systems remains a difficult challenge. Oversampling, needed to cover the expanded bandwidth of the predistortion signal, means that conventional DPD implementations would need digital circuits and data converters supporting bandwidths in the gigahertz in order to support ultra-wideband modulation signals for 5G. These requirements render the hardware implementation of real-time massive

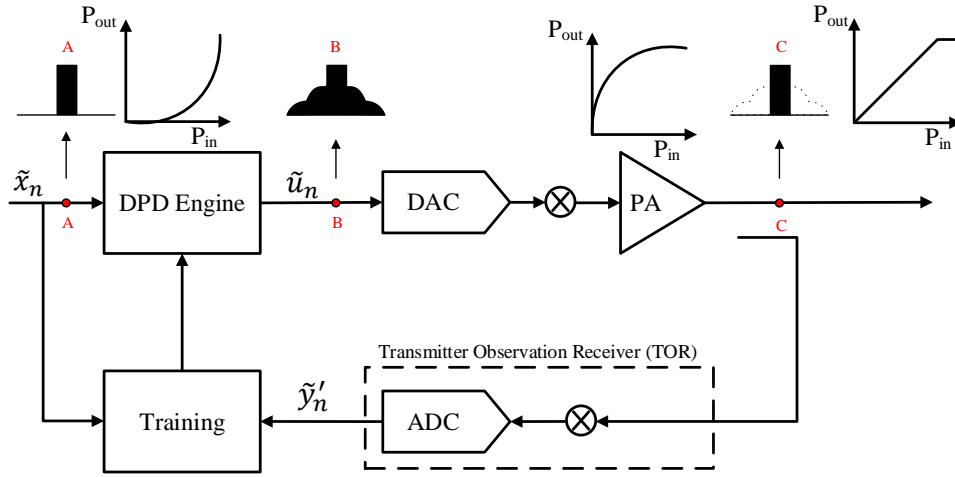


Figure 4.1: Block diagram of a typical DPD system.

MIMO DPD difficult, and ultimately can result in major energy and cost overhead in the underlying hardware. This chapter aims to address such predistortion engine implementation problems and presents a novel scalable parallel-processing-based DPD engine that dissociates the maximum linearization bandwidth from the maximum clock frequency of the digital circuit, thus enabling linearization of ultra-wideband signals with conventional digital circuits.

4.2 Parallel-Processing-Based Digital Predistortion Engine

Fig. 4.1 shows the architecture of a typical DPD system, consisting of a predistortion engine, a TOR and a training module. The DPD engine processes the complex baseband equivalent form of the original signal \tilde{x}_n , and computes the desired predistorted signal \tilde{u}_n . High-speed DACs convert the predistorted signal to the analog domain, and it is then up-converted to mm-wave frequency. For adaptive control, the PA output $y(t)$ is captured by the TOR, which consists of a down-converter stage followed by high-speed ADCs. The sampled output \tilde{y}_n is then used in the training module to update the coefficients for the DPD engine. The effect of PA nonlinearity creates spectral regrowth that occupies wider bandwidth. The inverse function of the PA nonlinearity generated by the DPD engine is also a nonlinearity function. The predistorted signal \tilde{u}_n required to cancel the PA distortion

also occupies wider bandwidth, usually 3-5x the bandwidth of the original signal. For example, to support wideband signals with modulation bandwidths of 800 MHz, both the DPD engine and the DACs need to allow complex data occupying 2.4–4 GHz of bandwidth.

Despite the availability of high-speed DACs operating at multiple giga-samples per second (GSPS), the DPD engine becomes a computational bottleneck due to its serial processing architecture. Many modern-day DPD engines are based on pruned Volterra series models that can be described as linear combinations of nonlinear basis functions with memory in the form

$$\tilde{u}_n = \sum_{m=0}^M a_{k,m} \psi_{n+m}^m \quad (4.1)$$

$$\psi_{n+m}^m = \sum_{k \in K} f_k(\tilde{x}_n, \tilde{x}_{n+1}, \dots, \tilde{x}_{n+m}) \quad (4.2)$$

where \tilde{x}_n and \tilde{u}_n are the n th samples of the input and predistorted signals respectively, M is the memory depth of the DPD engine, ψ_{n+m}^m is the collection of K basis with maximum memory m , and $a_{k,m}$, $k \in K, m = 1, \dots, M$ are the DPD engine coefficients. A typical real-time implementation of such a DPD engine would process one sample of the input signal per clock cycle and generate one sample of the predistorted signal per clock cycle. To optimize hardware resource usage, the DPD engine can be divided into two parts: the basis generation stage and the coefficient application stage. As illustrated in Fig. 4.2, assuming a memory depth of $M = 5$, the basis generation stage computes the value of each basis from the input samples, with memory basis such as ψ_{n+m}^m computed after the x_{n+m} has been consumed at the input. After inserting the buffers necessary for delay matching, the basis values ψ^0 to ψ^m from each memory branch are aligned together and fed to the coefficient application stage. Here, each of them is multiplied by a corresponding DPD coefficient and their summation yields u_n . Therefore, the maximum data throughput of this architecture is limited by the maximum clock rate of the underlying digital processing unit. Modern SOCs based on FPGAs, popular platforms for implementing DPD systems, typically have data paths with maximum clock frequencies in the range of several hundred megahertz. This is far below the required rate for wideband DPD systems. To achieve higher data throughput with the same data-path clock frequency, it is essential to apply the concept of parallelism and process multiple samples per clock cycle. Assuming a parallel acceleration factor of F at the input side, F samples of input signal will be streamed per clock cycle and F samples of the predistorted signal will need to be computed per clock cycle, for every clock cycle. This can be understood as F synchronous branches of interleaved samples with a sampling rate of $1/F$. It is apparent from (4.2) that simply replicating the serial DPD engine multiple times, one for each branch, would not work for

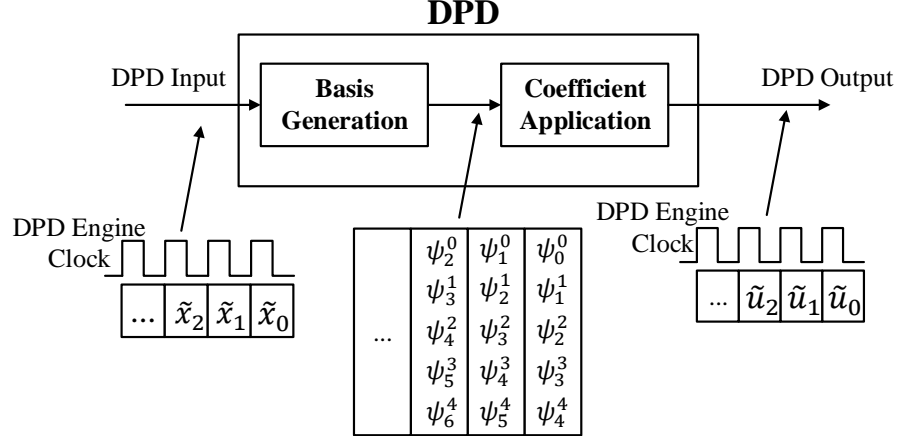


Figure 4.2: Data flow diagram of conventional serial processing DPD engine, assuming memory depth $M = 5$.

any DPD engine except a simple static model. For a DPD engine with memory depth M , each output sample u_n depends on the M input samples x_n to x_{n+M} , which are distributed in the F branches. Thus a new DPD engine architecture is required to support parallel processing.

To mitigate this issue, a new parallel processing DPD engine architecture is proposed, as shown in Fig. 4.3. Using a parallel acceleration factor of $F = 4$ and a memory depth of $M = 5$ as an example, the high-level data flow diagram of the proposed parallel processing DPD engine is shown in Fig. 4.4. The basis generation stage has been split into two steps. Each of the F samples of input signal is processed by a polynomial generation block with K output such that

$$\phi_n^k = f_k(\tilde{x}_n) \quad (4.3)$$

of each branch are computed, with the nonlinear polynomial function $f_k()$ usually taking the form of even order power series

$$f_k(\tilde{x}_n) = |\tilde{x}_n|^{2k} \quad (4.4)$$

for odd-order nonlinearities. The polynomial terms from all branches are then rearranged and cross multiplied in the cross-bar block to compute the correct basis ψ_{n+m}^m required for each sample of the predistorted signal u_n of the respective branch. The coefficient application stage for each output branch can then multiply the DPD coefficients with the aligned basis value, and sum the results for the correct output of u_n .

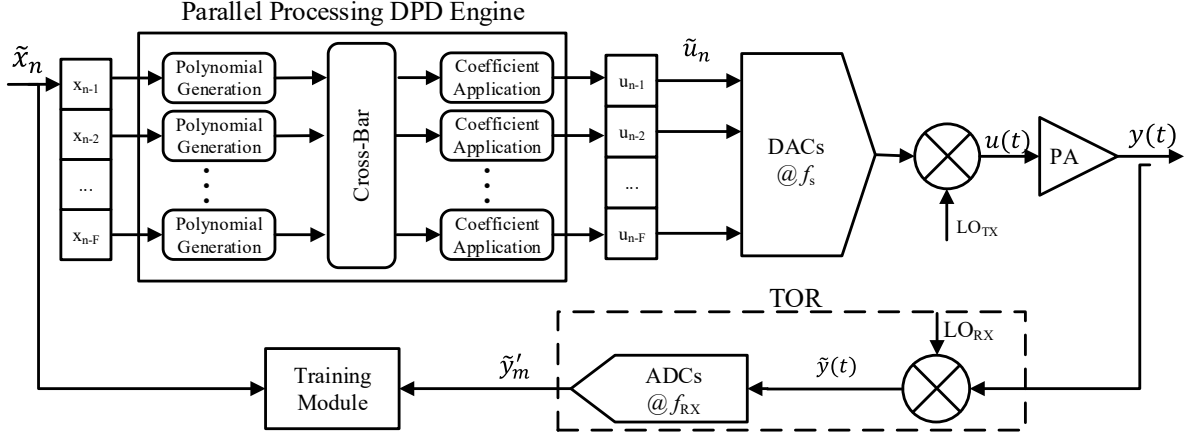


Figure 4.3: Block diagram of proposed system with parallel processing DPD engine, having parallel acceleration factor F and under-sampling feedback signal sampled at f_{RX} and sampled $\tilde{y}'_m = \tilde{y}(m \times R/f_s)$.

The detailed design and implementation of each block is tightly coupled with the design of the preceding and following stages, and is best illustrated with an example. Consider a DPD engine using the CRV model [43] as shown in (4.5), with non-linearity order $N = 9$, memory depth $M = 5$ and parallel acceleration factor $F = 4$.

$$\begin{aligned}
\tilde{u}_n = & \sum_{\substack{i=0 \\ \text{even}}}^{N-1} a_{1,i,0} \tilde{x}_n |\tilde{x}_n|^i \\
& + \sum_{m=1}^M \sum_{\substack{i=0 \\ \text{even}}}^{N-3} \sum_{\substack{k=2 \\ \text{even}}}^{N-1} a_{m,i,k} \tilde{x}_n |\tilde{x}_n|^i |\tilde{x}_{n-m}|^k \\
& + \sum_{m=1}^M \sum_{\substack{i=0 \\ \text{even}}}^{N-1} \sum_{\substack{k=2 \\ \text{even}}}^{N-1} b_{m,i,k} \tilde{x}_{n-m} |\tilde{x}_n|^i |\tilde{x}_{n-m}|^k
\end{aligned} \tag{4.5}$$

To ensure the resource and power efficiency of such an implementation, hardware optimization techniques should be applied where possible. Here, costly complex value multiplication can be delayed to latter stages by initially re-formulating the model to factor out the common terms \tilde{x}_n and \tilde{x}_{n-t} , as shown in (4.6).

$$\begin{aligned}
& \sum a_{1,i,0} \tilde{x}_n |\tilde{x}_n|^i \rightarrow \\
& \left(\sum \text{Re}(a_{1,i,0}) |\tilde{x}_n|^i + \text{Im}(a_{1,i,0}) |\tilde{x}_n|^i \right) \tilde{x}_n
\end{aligned} \tag{4.6}$$

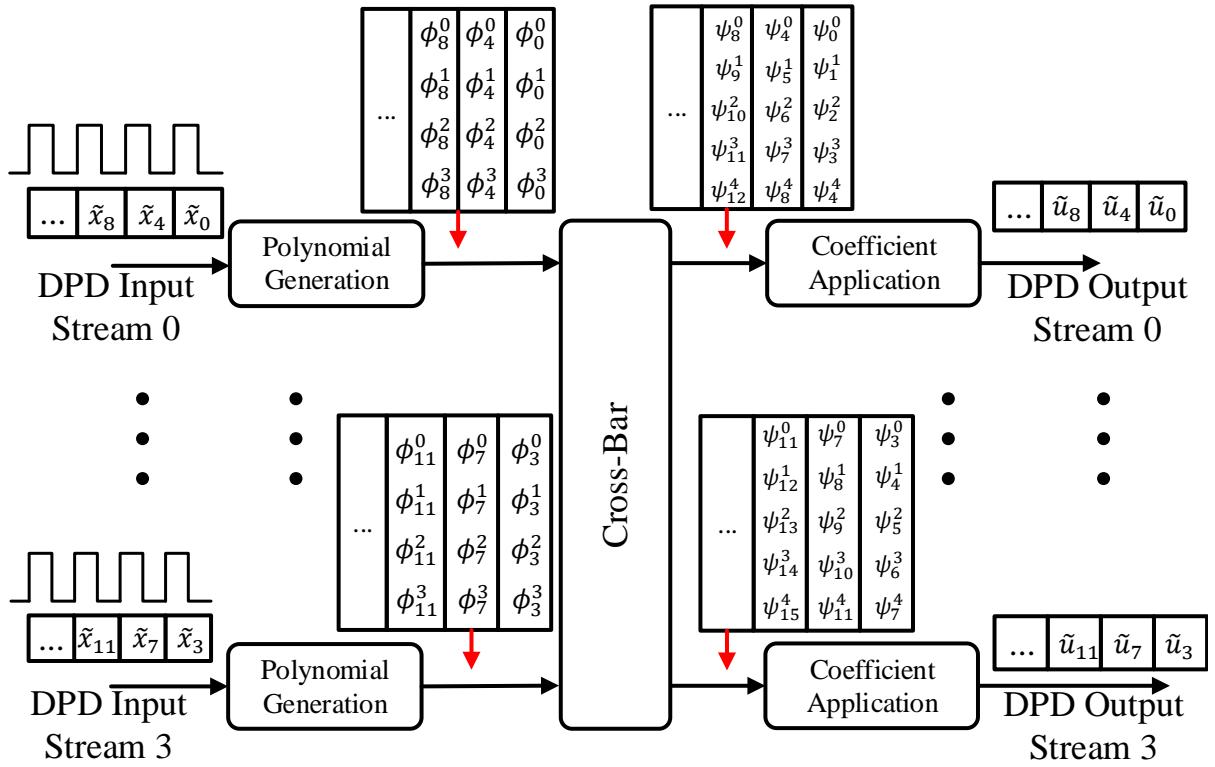


Figure 4.4: Data flow diagram of proposed parallel processing DPD engine, assuming memory depth $M = 5$ and parallel acceleration factor $F = 4$.

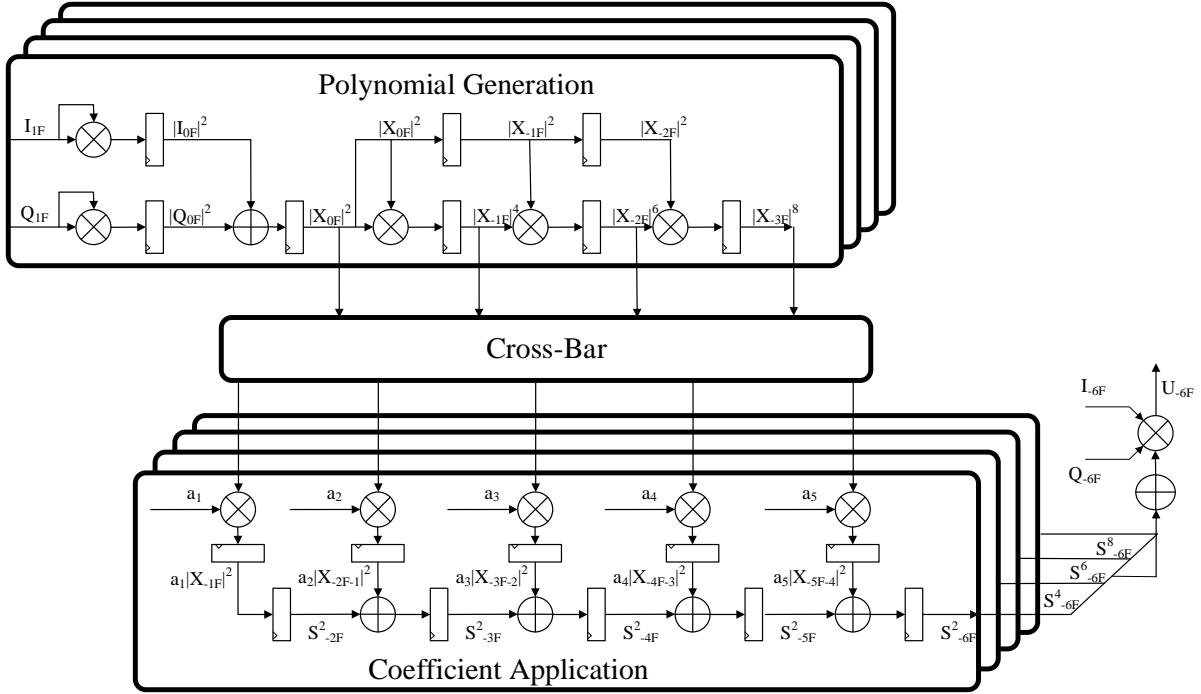


Figure 4.5: Detailed implementation of a section of the parallel-processing DPD engine illustrating the pipelined polynomial generation, the cross-bar, and one branch of the coefficient application stage.

This leaves only scalar-valued computation to be carried out on the envelope of the signal, described by the $|\tilde{x}_n|$ and $|\tilde{x}_{n-t}|$ terms. With F parallel data streams, the in-phase (I) and quadrature (Q) components of the signal, I_{nF+1} to $I_{(n+1)F}$ and Q_{nF+1} to $Q_{(n+1)F}$, arrive at the input of the DPD engine at each clock cycle. By following the standard practice of choosing only odd order nonlinear terms, the nonlinear polynomial terms of the envelope $|\tilde{x}_{nF+1}|^i$ to $|\tilde{x}_{(n+1)F}|^i$ will be even order. These even order polynomial terms are computed by first summing the squares of I and Q , and then raising them to the i th order in a pipeline fashion, as shown in Fig. 4.5.

The cross-bar block is placed after the polynomial generation blocks as it distributes the necessary terms across branches. The cross-bar block consists of a number of input buffer registers—one for every polynomial term from all branches—followed by a complicated routing network, as shown in Fig. 4.6. The input buffers are serial-in parallel-out shift registers of maximum depth $D = \text{ceil}((M + F)/F)$; there are a total of $(N - 1)/2 \times F$. There will be $N/2 \times M \times F$ output ports feeding the coefficient application stage of each

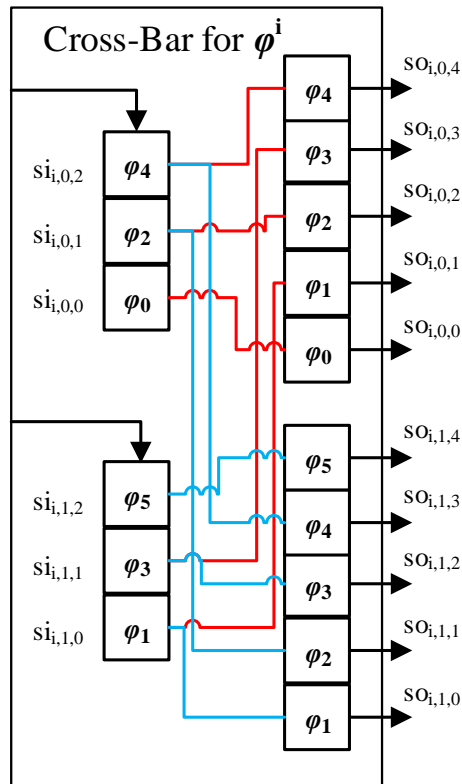


Figure 4.6: Block diagram of a sub-section of a simple cross-bar block with $F = 2$ and $M = 5$, with routing for each branch shown in different color.

branch. The routing between the input registers and the output ports is best carried out systematically, as follows. First, name each input register position $si_{i,f,d}$ and output port $so_{i,f,t}$ such that i is the power term order, f is the branch number, d is the depth inside the input register, and t is the memory depth required at the output port. The mapping from each $so_{i,f,t}$ to $si_{i,f,d}$ is therefore

$$i_{si} = i_{so} \quad (4.7)$$

$$f_{si} = \text{mod}(M - t_{so} + f_{so}, F) \quad (4.8)$$

$$d_{si} = \text{floor}((M - t_{so} + f_{so})/F) - M + (t_{so}) \quad (4.9)$$

Additional buffers may be inserted to account for pipeline latency.

After the cross-bar block has redistributed the power terms, the DPD basis with cross-terms such as $|\tilde{x}_n|^i |\tilde{x}_{n-t}|^k$ can now be computed from the power terms. From (4.5), it is apparent that different parts of the CRV model (i.e., $a_{t,i,k} \tilde{x}[n] |\tilde{x}[n]|^i |\tilde{x}[n-t]|^k$ and $b_{t,i,k} \tilde{x}[n-t] |\tilde{x}[n]|^i |\tilde{x}[n-t]|^k$), contain the same envelope cross terms and do not require additional computation. DPD coefficients are then applied to the envelope terms in a similar pipeline fashion, with the results summed to generate the complex predistorted signals, u_{nF+1} to $u_{(n+1)F}$, for all F branches.

Mm-wave PAs based on silicon processes have different distortion mechanisms than PAs based on III-V devices such as gallium nitride (commonly used for sub-6 GHz PAs). Based on the authors' observations of silicon-based PA behavior, it is hypothesized that the dynamic nonlinear memory effects in such PAs mainly depend on the envelope of the signal. This would allow them to be linearized by a DPD engine using an envelope only model such as the envelope complexity reduced Volterra (ECRV) model described in (4.10)

$$\begin{aligned} \tilde{u}_n = & \sum_{\substack{i=0 \\ \text{even}}}^{N-1} a_{1,i,0} \tilde{x}_n |\tilde{x}_n|^i \\ & + \sum_{t=1}^M \sum_{\substack{i=0 \\ \text{even}}}^{N-3} \sum_{\substack{k=2 \\ \text{even}}}^{N-1} a_{t,i,k} \tilde{x}_n |\tilde{x}_n|^i |\tilde{x}_{n-t}|^k \end{aligned} \quad (4.10)$$

This formulation has significantly fewer coefficients, meaning less hardware is used and less power consumed. Even though the ECRV model does not have any linear memory terms, such linear distortion can be treated as part of the channel and be equalized by the receiver.

Additional device-specific optimizations can be carried out at this stage. For example, on an FPGA platform, signals are often represented in fixed-point format for better

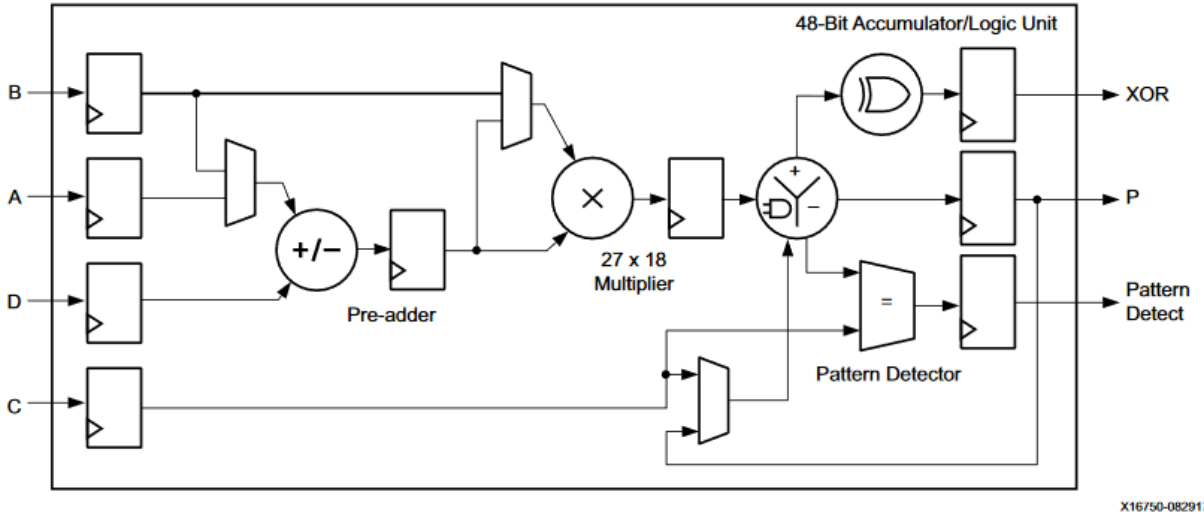


Figure 4.7: High level block diagram of DSP48 unit in Xilinx FPGA.

hardware resource utilization. With each multiplication, the number of bits of the signal increases. These bits must be truncated for the next operation. Many FPGAs have dedicated hardened DSP units that perform multiply-accumulate actions with asymmetric bus widths for each operand. For example, the DSP48 from Xilinx as shown in Fig. 4.7[104] has 18 and 27 bits for the two inputs to the multiplier and 48 bits for the accumulate action, and multiple stage of built-in registers for pipeline purpose. The DPD engine can take advantage of such features to optimize the signal quality along the data paths. Careful fixed-point simulation is required to evaluate the impact of such hardware operations in order to optimize the design to meet the requirement for adequate linearization.

4.3 Measurement Results

To assess the performance of the proposed real-time DPD system, two test cases were used under two different types of measurement setups. The first was a conducted measurement setup (i.e., measurements are performed via cables) using a silicon DPA as the device under test (DUT). The second was an OTA measurement setup with a 64-element beamforming array as the DUT. The DPD engine was implemented using the programmable logic of the FPGA on a Xilinx Ultrascale+ MPSOC evaluation board (ZCU102), and operated at a core clock rate of 300 MHz. The acceleration factor F was varied between 2 to 8 to achieve

scalable linearization bandwidth between 600 MHz to 2.4 GHz, corresponding to an over-sampling factor of 3 for the subsequent test signals occupying modulation bandwidths of 200–800 MHz. A pruned version of the CRV model in (4.5) was used for the DPD engine, with a nonlinearity order $N = 9$ and memory depth of $M = 5$ to yield a total of 43 complex coefficients. Additionally, an ECRV-based DPD engine [as described in (4.10)] was implemented, consisting of only 18 complex coefficients for the same nonlinearity order $N = 9$ and memory depth $M = 5$. The FPGA hardware resource utilization figures corresponding to both DPD engines, using the highest acceleration factor of $F = 8$, are summarized in Table. 4.1. The resulting dynamic power consumption from these implementations were estimated in Xilinx Vivado and are reported in Table. 4.1. As shown in Table 4.1, for the case of CRV with 43 coefficients, each branch of the DPD engine requires 104 DSP slices. In the case of the ECRV model, due to the greatly reduced number of coefficients, only 46 DSP slices are required per branch. This is a significant resource and power savings.

Fig. 4.8 shows the setup of the conducted measurements. The setup consists of a baseband stage, mm-wave up-converter and down-converter stages, and a DUT—a 28 GHz DPA fabricated in GlobalFoundries’ 45 nm SOI-CMOS process. The complex baseband input signals are represented in fixed point format with a resolution of 16 bits for both the I and Q components. According to Fig. 4.8, these baseband input signals are stored in the external DDR memory on the FPGA evaluation board and the memory interface is configured to read in F samples in parallel per clock cycle. Subsequently, the output of the DPD engine feeds the RF-DAC (AD9162 from Analog Devices) through a standard JESD204B high-speed converter interface. The RF-DAC runs at a clock rate of 4.8 GSPS

Table 4.1: Resource Utilization of Proposed DPD Engine (with nonlinearity order $N = 9$, memory depth $M = 5$ and parallel acceleration factor $F = 8$.)

	CRV		ECRV	
	Total	Per Branch	Total	Per Branch
No. Coef	43	43	18	18
DSP Slices	832	104	368	46
Slice LUTs	12138	1517	5369	671
Slice Registers	23038	2880	10190	1274
Memory	0	0	0	0
Power* (mW)	2166	271	958	120
Latency	23 clock cycle		20 clock cycle	

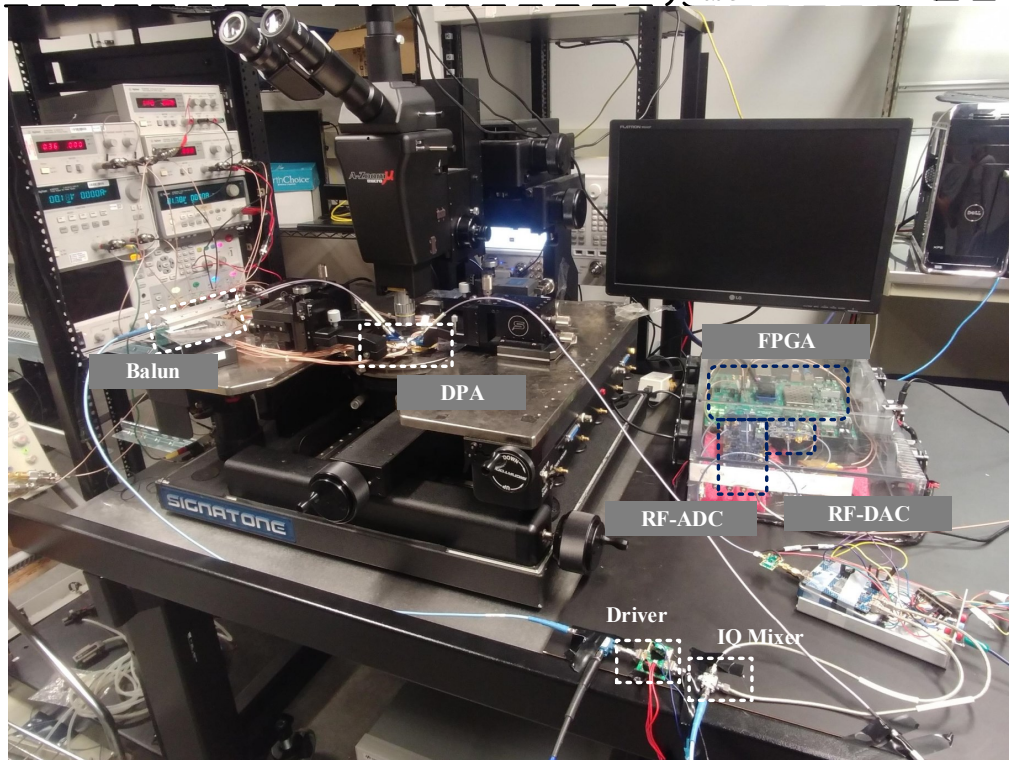
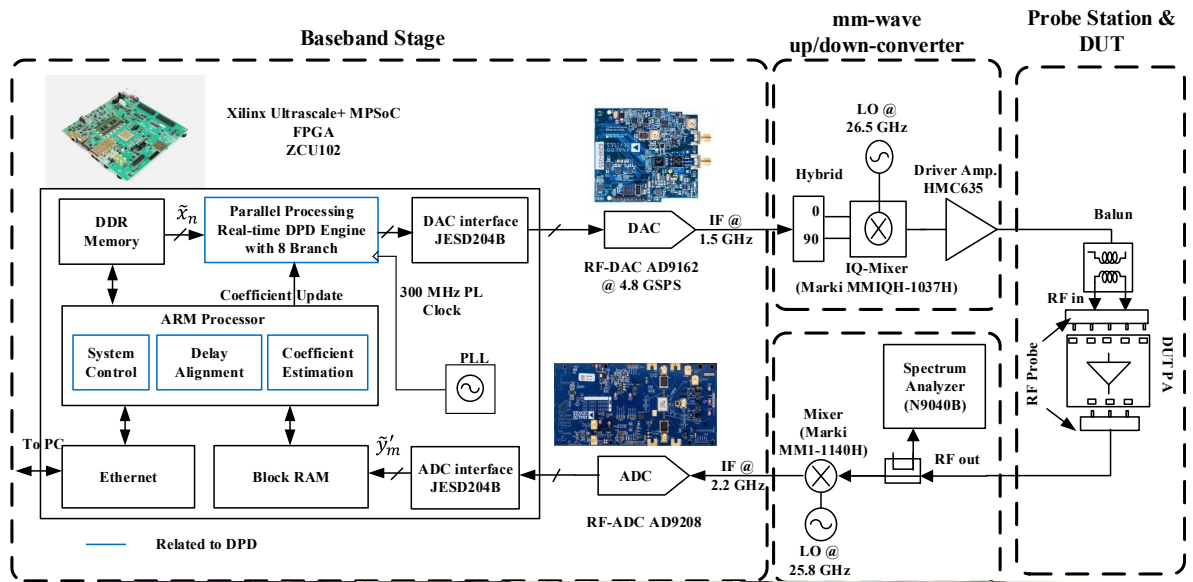


Figure 4.8: Measurement setup of proposed DPD system – linearizing a silicon DPA in a probe station.

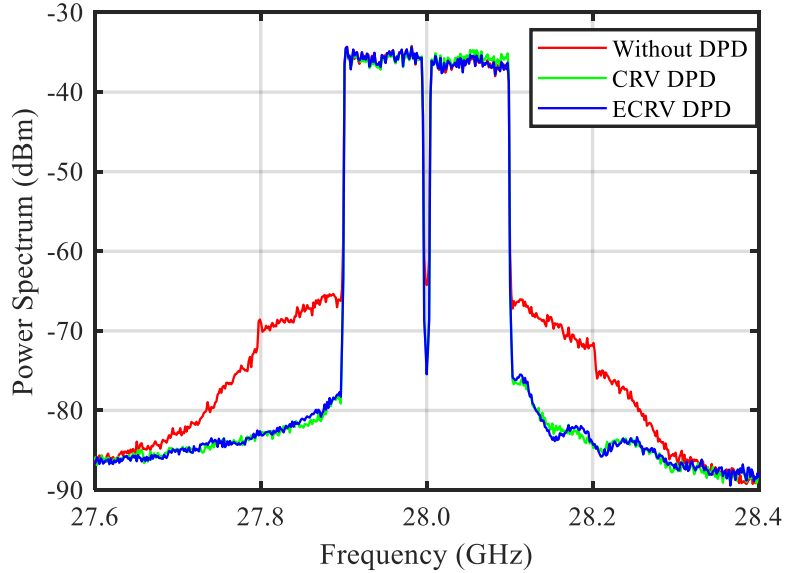


Figure 4.9: Output spectra of mm-wave silicon DPA in conducted measurement setup operating at 28 GHz using signals with 200 MHz bandwidths.

and uses a built-in digital-upconverter stage to translate the complex baseband signal to an IF frequency of 1.5 GHz. Then, the IF output from the RF-DAC is up-converted to 28 GHz using an image-rejection IQ up-converter followed by a driver amplifier before feeding to the DUT. The output of the DPA is down-converted to an IF frequency of 2.2 GHz and sampled. The digitized samples at the output of the RF-ADC feed the FPGA through a high-speed converter interface (JESD204B) and are subsequently stored in the on-chip block random access memory (RAM). The delay alignment and coefficient training are performed by the embedded ARM processor on the SOC, which has access to both the DDR memory (storing the original baseband waveforms) and the block RAM (storing the waveforms sampled by the TOR). The embedded processor is also responsible for coordinating different parts of the programmable logic. Operations on the FPGA board are monitored by the PC through an Ethernet connection.

Therefore, a spectrum analyzer (Keysight N9040B) is used to measure the ACPR and

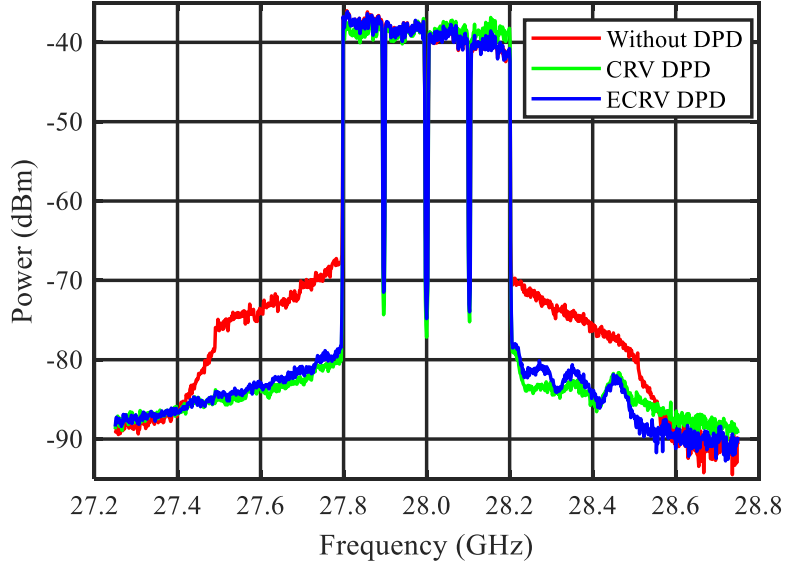


Figure 4.10: Output spectra of mm-wave silicon DPA in conducted measurement setup operating at 28 GHz using signals with 400 MHz bandwidths.

EVM. Here, EVM per OFDM symbol is computed as follows

$$\text{EVM} = \frac{\sum_{i=1}^{L_{DSC}} |\tilde{R}_{u(i)} S_{u(i)} - \tilde{X}_{u(i)}|^2}{\sum_{i=1}^{L_{DSC}} |\tilde{X}_{u(i)}|^2} \quad (4.11)$$

where \tilde{X}_u and \tilde{R}_u are the L_{FFT} -length discrete Fourier transform (DFT) of the input \tilde{x}_u and received \tilde{r}_u signals, respectively, L_{DSC} is the number of data subcarriers, $u(i)$ is the index of the DFT corresponding to the i -th data subcarrier, and S_u is a linear equalization filter derived from the OFDM pilot subcarriers. The test signals used in the experiments started with a minimum of 200 MHz modulation bandwidth formed by aggregating two component carriers (2-CC) of 100 MHz 64-QAM OFDM signals. A crest-factor reduction technique was applied to minimize the PAPR, yielding a PAPR of about 10 dB for the 2-CC test signals. Similarly, test signals of much wider bandwidths [4-CC (equivalently 400 MHz) and 8-CC (equivalently 800 MHz)] were constructed and tested with the proposed real-time DPD engine using acceleration factors of $F = 4$ and $F = 8$, respectively. Their PAPRs after crest factor reduction were about 10.5 dB.

Fig. 4.9 and Fig. 4.10 show the spectrum at the output of the DPA in the conducted measurements using 200 MHz and 400 MHz OFDM test signals, respectively. Their ACPR and EVM are summarized in Table 4.2. According to Table 4.2, for the case of 200 MHz test signals, a similar extent of linearization performance is observed using either the CRV- or ECRV-based DPD engines. The EVM reduced from 3.9% before DPD to 1.1% after DPD. The ACPR reduced from -32 dBc before DPD to around -46 dBc after DPD. In the case of the 400 MHz test signal, the EVM improvements were similar for both the CRV or ECRV engines (3.9% and 1.1%, respectively). Again, the CRV- and ECRV-based DPD engines achieved very similar ACPR improvements (-44.6 dBc and -44 dBc, respectively). Considering the significantly reduced number of coefficients (18 versus 43), the ECRV-based DPD engine showed very promising potential for 5G transmitters employing wideband OFDM signals.

To further assess the performance of the proposed real-time DPD system, OTA measurements were conducted to linearize a 64-element beamforming array (AWMF-0129) operating at 28 GHz. The block diagram of the OTA measurement setup is provided in Fig. 4.11. According to Fig. 4.11, the same baseband stage and up-converter highlighted in Fig. 4.8 are used to provide the mm-wave input to feed the array radio head placed inside the anechoic chamber. A receiving horn antenna was placed at the far-field of the main beam direction. The mm-wave signal captured by the receiving horn antenna was down-converted to an IF of 2.2 GHz. Similar to the conducted measurements, a spectrum analyzer was used to assess the performance metrics (ACPR and EVM) of the signal captured by the horn antenna. It is worth noting that significantly dispersive channel responses were noticed in the OTA measurements. This necessitated a calibration step using the procedures as described in [105, 88].

The output spectra of the mm-wave 64-element array captured by the OTA horn antenna are shown in Fig. 4.12 and Fig. 4.13 for the 400 MHz and 800 MHz test cases, respectively. The ACPR and EVM for both DPD engines are summarized in Table 4.2. According to Table 4.2, in the case of the 400 MHz test signals, both the CRV and ECRV engines demonstrated similar linearization capacity in terms of EVM improvements—from 3.9% before DPD to 1.1% after DPD. The CRV-based DPD engine achieved an ACPR improvement from -33 dBc to about -45.8 dBc, while the ECRV-based DPD engine showed about -44.2 dBc ACPR. In the case of the 800 MHz test signal, the CRV-based DPD engine achieved an EVM improvement from 4.1% to 1.1%, and an ACPR improvement from -33 dBc to -40 dBc after DPD. The ECRV-based DPD engine showed similar performances, with a post-DPD EVM of 1.2% and ACPR of -40 dBc.

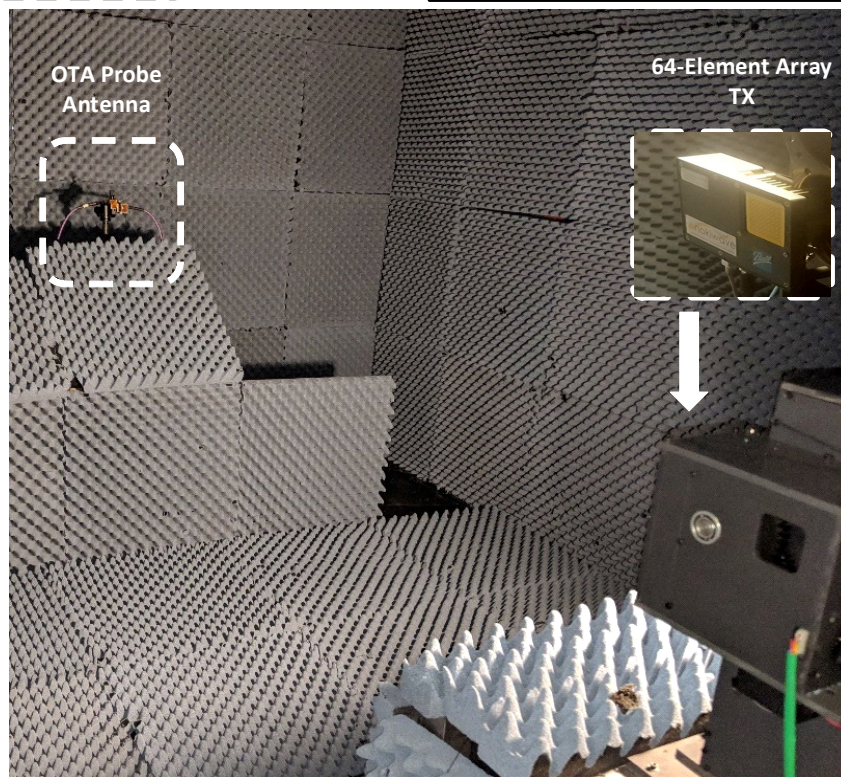
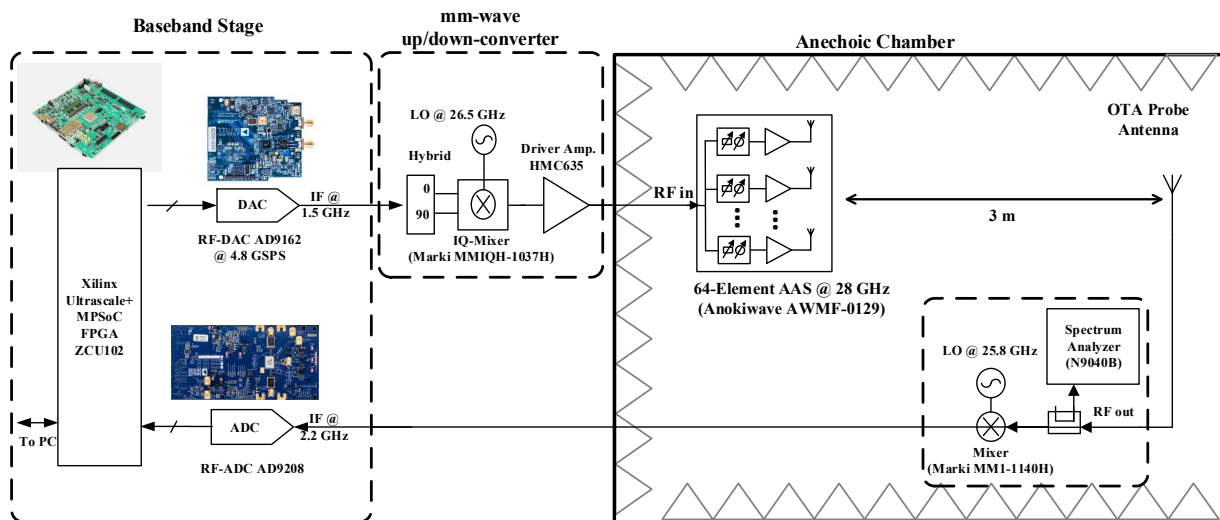


Figure 4.11: Measurement setup of proposed DPD system – linearizing a 64-element mm-wave beamforming array with OTA receiving antenna.

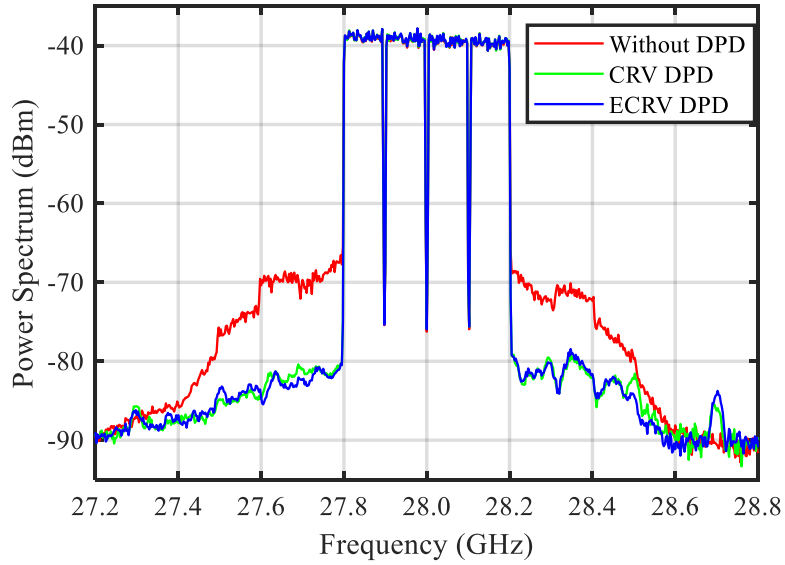


Figure 4.12: Output spectra of mm-wave 64-element beamforming array in the OTA measurement setup operating at 28 GHz using signals with 400 MHz bandwidth.

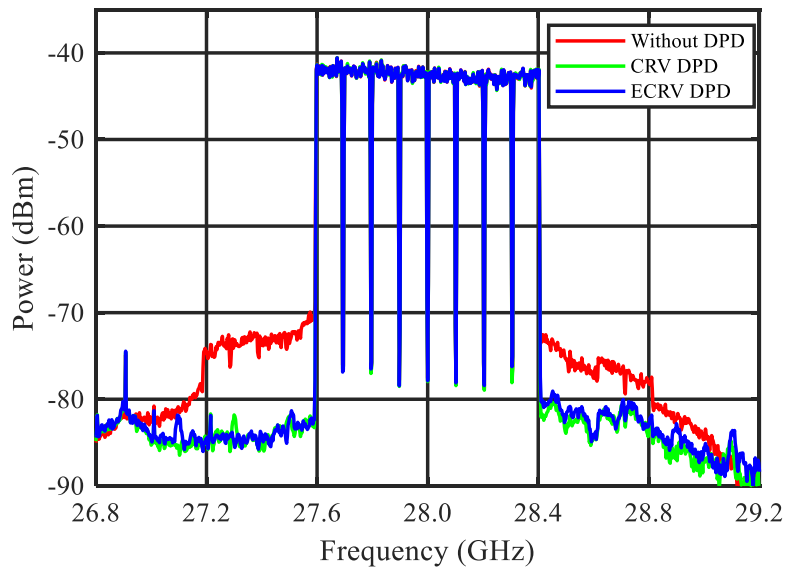


Figure 4.13: Output spectra of mm-wave 64-element beamforming array in the OTA measurement setup operating at 28 GHz using signals with 800 MHz bandwidth.

Table 4.2: Summary of Linearization Performance of mm-wave Silicon DPA and mm-wave 64-Element Beamforming Array (operating at 28 GHz).

	Conducted Meas. DPA (200MHz)		Conducted Meas. DPA (400MHz)		OTA Meas. Array (400MHz)		OTA Meas. Array (800MHz)	
	ACPR	EVM	ACPR	EVM	ACPR	EVM	ACPR	EVM
No DPD	-32.6/-34.9	3.9%	-32.6/-34.9	3.9%	-33.0/-35.0	3.9%	-33.6/-36.5	4.1%
CRV	-46.7/-46.7	1.1%	-44.6/-46.1	1.1%	-45.8/-46.5	1.1%	-40.7/-42.0	1.1%
ECRV	-46.4/-46.6	1.1%	-43.9/-45.5	1.1%	-44.6/-44.2	1.1%	-40.2/-40.9	1.2%

4.4 Conclusion

In this chapter, a hardware-efficient real-time DPD system with scalable linearization bandwidth for wideband sub-6 GHz and 5G mm-wave transmitters was presented. Using a novel parallel-processing DPD engine architecture, multiple samples can be processed per clock cycle, overcoming the maximum linearization bandwidth limit imposed by the maximum clock frequency of the digital circuits. The proposed DPD system is able to efficiently linearize wideband signals with potentially unlimited bandwidths. Hardware design choices that optimize resource usage are presented using a sample designed on a commercial FPGA. The proposed DPD system achieved over 2.4 GHz linearization bandwidth using only a 300 MHz core clock for the digital circuits. Using a silicon DPA with 200 MHz and 400 MHz signals as the DUT for conductive measurements, and a 64-element beamforming array with 400 MHz and 800 MHz signals as the DUT for OTA testing, the scalability and linearization capability of the proposed system were demonstrated at 28 GHz. Experimental results show that the DPD engine using an envelope-based model such as ECRV can achieve similar performance to conventional DPD models like CRV for linearizing a mm-wave DUTs transmitting wideband OFDM signals, while using less hardware and consuming less power.

Chapter 5

Predistortion Function Synthesis using Under-sampled Feedback Signals

5.1 Introduction

A modern predistortion system is not complete without a feedback path to observe the PA output and capture the distortion signal when updates to the linearization coefficients are needed. As discussed in previous chapters, intermodulation caused by nonlinear distortion of the PA results in expanded bandwidths of the PA output, typically 5 times the signal bandwidth. Conventional implementation of a TOR follows typical receiver topology and employs standard Nyquist rate sampling ADCs to capture the full bandwidth of the distortion signal. Such practices have become increasingly challenging to implement as the bandwidth of the signal increases. With 5G signals expected to increase to hundreds of megahertz, the design complexity and power consumption of the TOR becomes a serious issue. Not only does the ADCs in the TOR need to support extremely high sampling rates at several giga-samples per second, the digital interface and processing circuits following the ADC consume large amounts of hardware resources and power. For example, to support signals with modulation bandwidth of 800 MHz, the ADC needs a sampling rate of at least 8 GSPS, assuming an over-sampling factor of 5. The data throughput of the following digital interface and processing circuit reaches 96 Gb/s with 16-bit word length. A typical implementation of such an interface employing an advanced FPGA requires a complicated JESD204B interface protocol using 8 high-speed IO lanes with transceivers operating at

12 Gb/s, and the total power consumption of the converter interface alone consumes over 2 W of power.

To address this problem, this chapter presents an under-sampling framework that significantly reduces the sampling rate requirement of the ADC. It is based on the principle that a full reconstruction of the PA output signal is not required to characterize the PA distortion, which in turn allows sub-Nyquist sampling with significantly lower sampling speed. With an appropriate training algorithm, the coefficient updates can be computed using samples of the feedback signal captured by an ADCs running at a clock frequency significantly lower than the Nyquist rate. For clarity and simplicity, the discussion in this chapter is presented in the context of a DPD system, but the same principle can be used in an ARFPD system by replacing the digital operations with transfer functions of analog components such as multipliers.

5.2 Direct Learning Framework

Before presenting the under-sampling feedback, it is necessary to review the direct learning framework which form its basis. Briefly discussed in Chapter 2 Section 2.5.4, the direct learning architecture works on minimizing the difference between the original input and the normalized output. For most predistortion system using polynomial models, the general form of the model can be expressed as a summation of weighted basis with finite memory. With a DPD engine with memory depth M , and total number of basis K , the model can be expressed as

$$\tilde{u}_n = \sum_{m=0}^M a_{k,m} \psi_{n+m}^m \quad (5.1)$$

$$\psi_{n+m}^m = \sum_{k \in K} f_k(\tilde{x}_n, \tilde{x}_{n+1}, \dots, \tilde{x}_{n+m}) \quad (5.2)$$

where \tilde{x}_n and \tilde{u}_n are the n th samples of the input and predistorted signals respectively, M is the memory depth of the DPD engine, ψ_{n+m}^m is the collection of K basis with maximum memory m , and $a_{k,m}$, $k \in K, m = 1, \dots, M$ are the DPD engine coefficients. For a block of N samples $\tilde{x}_n, \dots, \tilde{x}_{n+N-1}$, (5.2) can be expressed in vector form

$$\mathbf{U} = \mathbf{\Psi} \mathbf{A}. \quad (5.3)$$

where \mathbf{U} is the N -length column vector of the predistorted signal $(\tilde{u}_n, \dots, \tilde{u}_{n+N-1})^T$, $\mathbf{\Psi} = [\Psi_{ij}]$ is a $N \times K$ matrix of the basis function with entries $\Psi_{ij} = \psi_{j,n+i-1}$, and \mathbf{A} is the

column vector of the K coefficients with $\mathbf{A} = (a_1, \dots, a_K)^T$. DPD models such as MP, GMP, EMP, DDR-Volterra discussed in Chapter 2 Section 2.5.3 can all be described in such way.

The PA is modelled as an unknown function $g(\cdot)$ with memory depth M_{PA} , hence:

$$\tilde{y}_n = g(\tilde{u}_n, \tilde{u}_{n+1}, \dots, \tilde{u}_{n+M_{PA}}), \quad (5.4)$$

where \tilde{y}_n is the sampled complex envelope of the output. With normalized input and output, the desired gain of the PA is unity, and the desired output is \tilde{x}_n . The error is defined as

$$\tilde{e}_n = \tilde{y}_n - \tilde{x}_n, \quad (5.5)$$

or in vector form

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}. \quad (5.6)$$

For a block of N samples, the cost function to be minimized for the residual error model is

$$J = E \{ |\tilde{e}_n|^2 \} = \sum_{n=1}^N \left| \tilde{e}_n - \sum_{j=1}^K \Delta a_j \psi_{j,n} \right|^2 \quad (5.7)$$

where Δa_i are the coefficients updates. Following the steepest descent method established in traditional LMS algorithm, the coefficient updates follows

$$\Delta \mathbf{A} = \mathbf{A}_{\ell+1} - \mathbf{A}_\ell = -\frac{1}{2} \mathbf{X} \frac{\partial \mathbf{J}_\ell}{\partial \mathbf{A}_\ell} = \mathbf{X} \mathbf{e}_\ell^* \frac{\partial \mathbf{Y}_\ell}{\partial \mathbf{A}_\ell}. \quad (5.8)$$

where \mathbf{A}_ℓ , \mathbf{X}_ℓ , \mathbf{Y}_ℓ , \mathbf{J}_ℓ are the coefficients, input and output signal blocks and the cost function values at ℓ th iteration respectively. Assuming the model is weakly nonlinear and $\Delta \mathbf{A}$ is small, we can approximate the error by

$$\mathbf{e} = \hat{\mathbf{e}} + \varepsilon \quad (5.9)$$

where $\hat{\mathbf{e}}$ is the vector of the modelled error defined as

$$\hat{\mathbf{e}} = \Psi \Delta \mathbf{A} \quad (5.10)$$

with ε being the residue. This eventually lead to the LSs solution

$$\Delta \mathbf{A} = (\Psi^H \Psi)^{-1} \Psi^H \mathbf{e}. \quad (5.11)$$

and an update factor γ for the LMS algorithm such that

$$\mathbf{A}_{\ell+1} = \mathbf{A}_{\ell} + \gamma \Delta \mathbf{A}. \quad (5.12)$$

with initial A_0 chosen such that the DPD engine initially passes through \tilde{x}_n undistorted, i.e., $\tilde{u}_n = \tilde{x}_n$.

In principle, Ψ could have as few as $N = K$ rows, the number of unknown coefficients in $\Delta \mathbf{A}$. However, due to quantization errors and ill-conditioning of the matrix $\Psi^H \Psi$ when N is small, better accuracy can be obtained by choosing $N > K$. Thus, the matrix operations above require significant computational complexity. A simpler method to estimate the residual error vector coefficients $\Delta \mathbf{A}$ is to use the well-known RLS algorithm, which updates the $\Delta \mathbf{A}$

$$\begin{aligned} \Delta \mathbf{A}_n &= \Delta \mathbf{A}_{n-1} + \mathbf{k}_n e_n - \Psi_n \Delta \mathbf{A}_{n-1} \\ \mathbf{k}_n &= \frac{\mathbf{P}_{n-1} \Psi_n^H}{1 + \Psi_n \mathbf{P}_{n-1} \Psi_n^H} \\ \mathbf{P}_n &= \mathbf{P}_{n-1} - \mathbf{k}_n \Psi_n \mathbf{P}_{n-1} \end{aligned} \quad (5.13)$$

starting with $\Delta \mathbf{A}_0 = \mathbf{0}$, ending with $\Delta \mathbf{A} = \Delta \mathbf{A}_N$, where Ψ_n is the row vector $(\psi_{1,n}, \dots, \psi_{K,n})$.

5.3 DPD Function Synthesis Using Under-sampled Feedback Signal

To reduce the hardware resource usage and power consumption of the feedback path, the under-sampling framework suggests the ADC in the feedback path to operate at a clock rate that is a fraction $1/P$ of the full-rate f_s , as illustrated in Fig. 5.1 The ADC must still have wide analog input bandwidth that is sufficient to capture the full expanded bandwidth of the distorted signal. While sampling the signal at significant lower rate than the Nyquist bandwidth, aliasing will happen. The PA output signal cannot be reconstructed as in the case of conventional TOR with Nyquist rate. However, contrary to the indirect learning architecture which relies on complete knowledge of PA outputs to construct the post-inverse, the direct learning architecture computes the coefficients of the pre-inverse directly using the input signal and PA output. This key difference is important as direct learning can still work even with incomplete information of the PA output (i.e. with sub-Nyquist sampling).

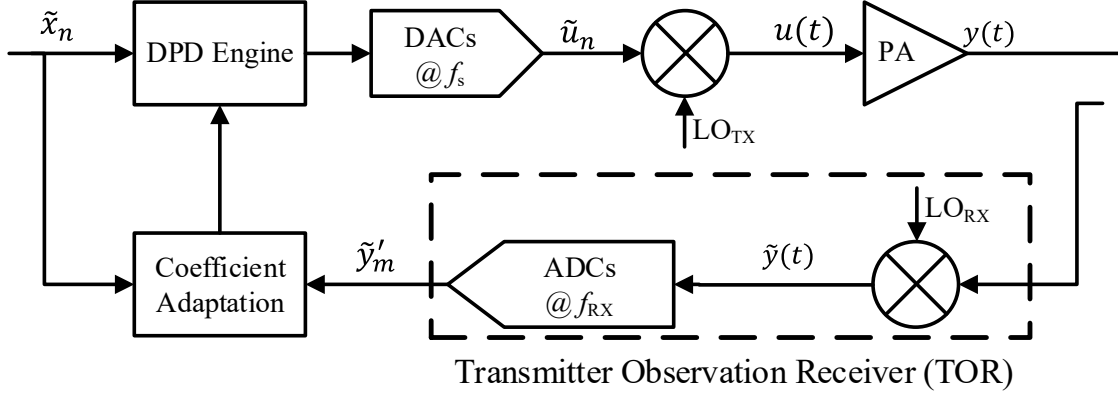


Figure 5.1: Block diagram of the proposed under-sampling feedback system with a feedback signal sampled at significantly reduced rate of f_{RX} . The conventional system corresponds to the special case where $f_{RX} = f_s$.

To illustrate this, a close look at (5.10) and expanded in matrix-vector form

$$\begin{bmatrix} \tilde{e}_n \\ \tilde{e}_{n+1} \\ \tilde{e}_{n+2} \\ \vdots \\ \tilde{e}_{n+N} \end{bmatrix} = \begin{bmatrix} \psi_{1,n} & \psi_{2,n} & \cdots & \psi_{K,n} \\ \psi_{1,n+1} & \psi_{2,n+1} & \cdots & \psi_{K,n+1} \\ \psi_{1,n+2} & \psi_{2,n+2} & \cdots & \psi_{K,n+2} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1,n+N} & \psi_{2,n+N} & \cdots & \psi_{K,n+N} \end{bmatrix} \begin{bmatrix} \Delta a_1 \\ \Delta a_2 \\ \Delta a_3 \\ \vdots \\ \Delta a_K \end{bmatrix}. \quad (5.14)$$

It is quite apparent that the matrix equation still holds with reduced sampling rate in the signals, with the same set of coefficients. For example, if only every other samples of the input and output are used, (5.14) is changed to

$$\begin{bmatrix} \tilde{e}_n \\ \tilde{e}_{n+2} \\ \tilde{e}_{n+4} \\ \vdots \\ \tilde{e}_{n+N} \end{bmatrix} = \begin{bmatrix} \psi_{1,n} & \psi_{2,n} & \cdots & \psi_{K,n} \\ \psi_{1,n+2} & \psi_{2,n+2} & \cdots & \psi_{K,n+2} \\ \psi_{1,n+4} & \psi_{2,n+4} & \cdots & \psi_{K,n+4} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{1,n+N} & \psi_{2,n+N} & \cdots & \psi_{K,n+N} \end{bmatrix} \begin{bmatrix} \Delta a_1 \\ \Delta a_2 \\ \Delta a_3 \\ \vdots \\ \Delta a_K \end{bmatrix}. \quad (5.15)$$

To put it formally, if the ADC is sampling at a clock rate that is a fraction $1/P$ of the full-rate f_s , the under-sampled output from the ADC is

$$\tilde{y}'_m = \tilde{y}(m \times T_s P) = \tilde{y}_n|_{n=PM}. \quad (5.16)$$

The associated error is

$$\tilde{e}'_m = \tilde{y}_m - \tilde{x}_{Pm}, \quad (5.17)$$

which can be approximated in the similar manner as (5.10) as

$$\hat{\mathbf{e}}' = \mathbf{\Psi}' \Delta \mathbf{A} \quad (5.18)$$

with the updated base matrix $\mathbf{\Psi}'$ been the reduced ranked matrix from $\mathbf{\Psi}$ which

$$\psi'_{j,m} = \psi_{j,Pm}. \quad (5.19)$$

The cost function for this under-sampled signals can be formulated in

$$J' = E \left\{ |\tilde{e}'_m|^2 \right\} = \sum_{n=1}^N \left| \tilde{e}_{nP} - \sum_{j=1}^K \Delta a_j \psi_{j,nP} \right|^2 \quad (5.20)$$

which corresponds to only including every P th terms from (5.7). Provided enough e'_m have been collected, the optimal $\Delta a'_i$ in (5.20) are close to those of (5.7) as they model the same error. Following the same derivation as in (5.11), the LSs solution of the coefficient updates is

$$\Delta \mathbf{A} = (\mathbf{\Psi}'^H \mathbf{\Psi}')^{-1} \mathbf{\Psi}'^H \mathbf{e}'. \quad (5.21)$$

The RLS algorithm minimizing J' is then

$$\begin{aligned} \Delta \mathbf{A}'_m &= \Delta \mathbf{A}'_{m-1} + \mathbf{k}'_m e'_m - \mathbf{\Psi}'_m \Delta \mathbf{A}'_{m-1} \\ \mathbf{k}'_m &= \frac{\mathbf{P}'_{m-1} \mathbf{\Psi}'_m{}^H}{1 + \mathbf{\Psi}'_m \mathbf{P}'_{m-1} \mathbf{\Psi}'_m{}^H} \\ \mathbf{P}'_m &= \mathbf{P}'_{m-1} - \mathbf{k}'_m \mathbf{\Psi}'_m \mathbf{P}'_{m-1} \end{aligned} \quad (5.22)$$

The same coefficient updates that lead to convergence of the predistortion coefficients can be obtained from under-sampled data, provided sufficient number of samples are collected to ensure good convergence towards the optimal coefficient values. This can be achieved by taking longer capture time or running more iteration loops.

5.4 DPD Function Synthesis Using Under-sampled Feedback Signal at IF

Many modern receiver architecture perform the IQ demodulation in digital domain by first down-convert the signal to low-IF and sampled by the ADC. This avoids the IQ imbalance problem presented in analog circuits, which only become worse with increasing signal

bandwidth. However, the under-sampling framework presented in previous section assumes the feedback signal to be sampled by the ADCs are at baseband, and is only suitable for applications where analog IQ demodulator is used to down-convert the signal to baseband in-phase and quadrature part. Though successful demonstration of the technique is presented in [85, 86], it cannot be applied directly to IF sampling receivers. As presented in [105], the aliasing effect of the sub-Nyquist sampling ADC is different for the baseband and IF sampling scheme. In the baseband sampling scheme, the pair of the ADCs captures the in-phase and quadrature components separately and the complex envelope of the baseband is aliased. However, in the IF sampling scheme, the real IF is captured by one under-sampling ADC and aliased. The algorithm presented in previous section cannot be used, and a new solution needs to be devised.

In conventional IF sampling receivers, assuming the complex envelope of the signal has bandwidth of f_s , the IF signal y_{IF} needs to be sampled by an ADC running at a clock rate f_{RX} of at least $2f_s$ to avoid aliasing. Assuming the complex envelope of the signal at baseband is \tilde{x}_n and \tilde{y}_n , the sampled IF signal can be expressed as

$$y_{n,IF} = \text{Re} \left\{ \tilde{y}(t) e^{j2\pi(f_{IF}t + \theta)} \right\} \Big|_{t=n/2f_s} \quad (5.23)$$

$$= \text{Re} \left\{ \tilde{y}_n e^{j2\pi n(f_{IF}/f_{RX}) + j2\pi\theta} \right\} \quad (5.24)$$

where $\tilde{y}(t)$ is the continuous time envelope of the baseband signal, θ is the phase of the IF carrier at $t = 0$ and Re is the real part operator. Without loss of generality, the phase of the carrier is assumed to be 0 and will be dropped for the rest of the derivation for compactness. Following the direct learning framework in the previous sections, the IF under-sampling work aims to minimize the error at the IF signal. The error of the IF signal is defined as

$$e_{n,IF} = \text{Re} \{ \tilde{e}_n \} \quad (5.25)$$

$$= \text{Re} \{ \tilde{y}_n \} - \text{Re} \{ \tilde{x}_n \} \quad (5.26)$$

$$= y_{n,IF} - \text{Re} \left\{ \tilde{x}_n e^{j2\pi n(f_{IF}/f_{RX})} \right\}. \quad (5.27)$$

Let x_{IF} be the up-converted IF version of the baseband input signal \tilde{x} such that

$$x_{n,IF} = \text{Re} \left\{ \tilde{x}_n e^{j2\pi n(f_{IF}/f_{RX})} \right\}, \quad (5.28)$$

(5.27) can be expressed as

$$e_{n,IF} = y_{n,IF} - x_{n,IF}. \quad (5.29)$$

Using the same weakly nonlinear assumption and assuming $\Delta\mathbf{A}$ is small, the IF error can be approximate by

$$\mathbf{e}_{IF} = \hat{\mathbf{e}}_{IF} + \varepsilon_{IF} \quad (5.30)$$

where $\hat{\mathbf{e}}_{IF}$ is the column vector of the approximated error and ε_{IF} the residue. $\hat{\mathbf{e}}_{IF}$ can be derived from (5.10) and (5.30) as

$$\hat{\mathbf{e}}_{IF} = \text{Re} \{ \hat{\mathbf{e}} e^{j2\pi n(f_{IF}/f_{RX})} \} \quad (5.31)$$

$$= \text{Re} \{ \Psi \Delta\mathbf{A} e^{j2\pi n(f_{IF}/f_{RX})} \} \quad (5.32)$$

$$= \text{Re} \{ \Delta\mathbf{A} \} \text{Re} \{ \Psi e^{j2\pi n(f_{IF}/f_{RX})} \} - \text{Im} \{ \Delta\mathbf{A} \} \text{Im} \{ \Psi e^{j2\pi n(f_{IF}/f_{RX})} \} \quad (5.33)$$

with ε_{IF} is the residue. Rearrange (5.33) in matrix form

$$\hat{\mathbf{e}}_{IF} = \begin{bmatrix} \text{Re} \{ \Psi e^{j2\pi n(f_{IF}/f_{RX})} \} \\ - \text{Im} \{ \Psi e^{j2\pi n(f_{IF}/f_{RX})} \} \end{bmatrix} \begin{bmatrix} \text{Re} \{ \Delta\mathbf{A} \} \\ \text{Im} \{ \Delta\mathbf{A} \} \end{bmatrix} \quad (5.34)$$

Let

$$\Psi_{IF} = \begin{bmatrix} \text{Re} \{ \Psi e^{j2\pi n(f_{IF}/f_{RX})} \} \\ - \text{Im} \{ \Psi e^{j2\pi n(f_{IF}/f_{RX})} \} \end{bmatrix} \quad (5.35)$$

$$\Delta\mathbf{A}_{IF} = \begin{bmatrix} \text{Re} \{ \Delta\mathbf{A} \} \\ \text{Im} \{ \Delta\mathbf{A} \} \end{bmatrix} \quad (5.36)$$

(5.34) becomes

$$\hat{\mathbf{e}}_{IF} = \Psi_{IF} \Delta\mathbf{A}_{IF} \quad (5.37)$$

which has a similar form as (5.10). Therefore, by minimizing the IF cost function

$$J_{IF} = E \{ |\tilde{e}_{IF}|^2 \} \quad (5.38)$$

the coefficient update can be calculated as

$$\Delta\mathbf{A}_{IF} = (\Psi_{IF}^H \Psi_{IF})^{-1} \Psi_{IF}^H \mathbf{e}_{IF}. \quad (5.39)$$

The under-sampling framework from the previous section can now be applied to treat the aliased real IF signal in a similar manner. The under-sampled output of the ADC under the IF scheme is now

$$\tilde{y}'_{m,IF} = \tilde{y}_{IF}(m \times T_s P) = \tilde{y}_{n,IF} \Big|_{n=Pm}. \quad (5.40)$$

The associated error is

$$\tilde{e}'_{m,IF} = \tilde{y}_{m,IF} - \tilde{x}_{Pm,IF}, \quad (5.41)$$

which can be approximated in the similar manner as (5.37) as

$$\hat{\mathbf{e}}'_{IF} = \mathbf{\Psi}'_{IF} \Delta \mathbf{A}_{IF} \quad (5.42)$$

with the updated base matrix $\mathbf{\Psi}'_{IF}$ been the reduced ranked matrix from $\mathbf{\Psi}_{IF}$ which

$$\psi'_{j,m,IF} = \psi_{j,Pm,IF}. \quad (5.43)$$

The cost function used for the under-sampled signals is

$$J'_{IF} = E \left\{ |\tilde{e}'_{m,IF}|^2 \right\} \quad (5.44)$$

And the LSs solution of the coefficient updates is

$$\Delta \mathbf{A}_{IF} = (\mathbf{\Psi}'_{IF}{}^H \mathbf{\Psi}'_{IF})^{-1} \mathbf{\Psi}'_{IF}{}^H \mathbf{e}'_{IF}. \quad (5.45)$$

5.4.1 Hardware Optimization of Under-sampled Feedback

Though the theoretical framework presented is general to all IF frequency, in practical implementation the choice of IF and TOR ADC sampling rate has significant impact on the complexity of the hardware. Certain choice of the IF f_{IF} and ADC sampling rate f_{RX} can result in reduced resource usage by taking advantage of special relationships between the selected frequencies.

The IF frequency f_{IF} is selected to be

$$f_{IF} = f_{RX} \frac{R}{4} \quad (5.46)$$

where R is an integer that allows f_{IF} to fall in the correct range for the down-converter. This particular selection of IF frequency results in

$$\tilde{y}'_m = \text{Re} \left\{ \tilde{y}(t) e^{j2\pi(f_{IF}t + \theta)} \right\} \Big|_{t = \frac{m}{f_{RX}} = \frac{mP}{f_s}} \quad (5.47)$$

$$= \text{Re} \left\{ \tilde{y}(mP/f_s) e^{j2\pi(mf_{IF}/f_{RX})} e^{j2\pi\theta} \right\} \quad (5.48)$$

$$= \text{Re} \left\{ \tilde{y}_{mP} e^{j2\pi(\frac{mR}{4})} e^{j2\pi\theta} \right\} \quad (5.49)$$

$$= \text{Re} \left\{ (j)^l \tilde{y}_n e^{j2\pi\theta} \right\} \Big|_{n=mP, l=mR/4} \quad (5.50)$$

Based on (5.50), and depending on the choice of R , the output \tilde{y}'_m of the under-sampling ADC will be reduced to either the in-phase $\pm \text{Re}\{\tilde{y}_n e^{j2\pi\theta}\}$ or quadrature component

$\pm \text{Im}\{\tilde{y}_n e^{j2\pi\theta}\}$ of $\tilde{y}_n e^{j2\pi\theta}$ (i.e., the full sample of \tilde{y}_n rotated by the constant phase θ). Such constant phase shift is linear with respect to the original signal and is tolerated in the DPD system.

Following the under-sampling framework presented previously, Let

$$q = mR/4 \quad (5.51)$$

the error for each sample is now

$$\tilde{e}'_n = \begin{cases} \tilde{y}'_m \pm \text{Re}\{\tilde{x}'_{mP}\} & \text{For } q \text{ even} \\ j(\tilde{y}'_m \pm \text{Im}\{\tilde{x}'_{mP}\}) & \text{For } q \text{ odd} \end{cases} \quad (5.52)$$

The approximated error terms become \hat{e}'_n are the modelling errors defined by

$$\hat{e}'_n = \begin{cases} \text{Re}\{\Psi'_n\} \text{Re}\{\Delta\mathbf{A}\} - \text{Im}\{\Psi'_n\} \text{Im}\{\Delta\mathbf{A}\} \\ \text{Im}\{\Psi'_n\} \text{Re}\{\Delta\mathbf{A}\} + \text{Re}\{\Psi'_n\} \text{Im}\{\Delta\mathbf{A}\} \end{cases} \quad (5.53)$$

The least squares estimate solution to this function is

$$\begin{bmatrix} \text{Re}\{\Delta\mathbf{A}\} \\ \text{Im}\{\Delta\mathbf{A}\} \end{bmatrix} = \begin{cases} (\Psi'_{In})^H \Psi'_{In})^{-1} \Psi'_{In}{}^H \tilde{e}'_n & \text{For } q \text{ even} \\ (\Psi'_{Qn})^H \Psi'_{Qn})^{-1} \Psi'_{Qn}{}^H \tilde{e}'_n & \text{For } q \text{ odd} \end{cases} \quad (5.54)$$

where

$$\Psi'_{In} = [\text{Re}\{\Psi'_n\} \quad -\text{Im}\{\Psi'_n\}] \quad (5.55)$$

$$\Psi'_{Qn} = [\text{Im}\{\Psi'_n\} \quad +\text{Re}\{\Psi'_n\}] \quad (5.56)$$

The DPD coefficients are then updated after sufficient samples have been collected for a good estimation of $\Delta\mathbf{A}$

$$\mathbf{A}_{w+1} = \mathbf{A}_w + \gamma \Delta\mathbf{A} \quad (5.57)$$

5.5 Delay Alignment

Accurate alignment of data is critical for direct learning. With the output of the PA sampled at a significantly lower rate, the delay alignment between \tilde{y}'_m and \tilde{x}_n becomes a challenge. To reduce complexity and improve robustness, two step delay estimation and compensation is performed.

The under-sampled signal \tilde{y}'_m is first up-sampled to full-rate $\tilde{y}'_{n,FR}$ by inserting zeroes between the samples. No further filtering to limit the bandwidth is performed, contrary to typical resampling algorithm. The integer delay with respect to the full-rate input signal is first estimated from the cross-covariance of small sections of the envelope

$$C(w) = \sum_m (|\tilde{x}_{n+p}| - \mu_x) (|\tilde{y}'_{n,FR}| - \mu_y)^*, \quad (5.58)$$

where μ_x and μ_y are the sample means of the small sections.

If the optimized scheme in Section 5.4.1 is used, the delay alignment has been made easier with the selection of the IF frequency, as each sample of the ADC output \tilde{y}'_m directly corresponds to either the I or Q of every P th sample in \tilde{x}_n . And the computationally expensive operation of complex envelope is reduced to a simple operation of reading the real or imaginary parts of the complex signal. The cross-covariance is calculated as

$$C(w) = \sum_m \begin{cases} \text{Re} \{ \tilde{x}_{mP+w} \} \tilde{y}'_m & \text{For } l \text{ even} \\ j \text{Im} \{ \tilde{x}_{mP+w} \} \tilde{y}'_m & \text{For } l \text{ odd} \end{cases} \quad (5.59)$$

The integer delay is estimated as

$$\hat{w} = \arg \max_w C(w). \quad (5.60)$$

After the integer delay has been aligned, the remaining fractional sample delay $\hat{d} \in (-1, 1)$ with respect to \hat{w} is then found by applying the fractional delay filter to sections of \tilde{x}_n and searching for the maximum value of $C(w + \hat{d})$. The fractional delay filter with delay \hat{d} is constructed from windowed sinc as

$$h[n] = W[n] \text{sinc}[n - \hat{d}] \quad (5.61)$$

where $W[n]$ is a window function such as Kaiser or Chebyshev, and the order of the filter is $2L + 1$. The choice of the window function and the length of the filter is determined based on the requirements of the system. After delay compensation of the input samples, the direct learning algorithm described in the previous section is used to estimate the DPD coefficients.

Table 5.1: Measurement Results of Baseband Based Training with Under-sampled Feedback Signal

	20 MHz Signal		80 MHz Signal	
	Sampling Rate (complex)	ACLR L/U	Sampling Rate (complex)	ACLR L/U
No DPD	—	-32.6/-31.4	—	-38.2/-30.9
Indirect	100 MSPs	-54.1/-54.9	400 MSPs	-48.6/-48.9
Full-rate	100 MSPs	-54.8/-54.8	400 MSPs	-50.5/-49.7
$N_u = 2$	50 MSPs	-55.2/-55.2	200 MSPs	-49.7/-49.4
$N_u = 5$	20 MSPs	-55.2/-54.9	80 MSPs	-49.8/-48.8

5.6 Validation and Measurement Results

5.6.1 Under-sampling using Baseband Scheme

The performance of the new approach using the baseband scheme is assessed experimentally to linearize a 20 W GaN DPA. The full-rate output of the DPD engine is computed from the input and an AWG is used to generate the RF signal driving a GaN PA operating at 2 GHz. The block diagram of the measurement setup is shown in Fig. 5.2. A wideband down-converter and low-pass filter is used to acquire the envelope of the PA output, which is then sampled. A 20 MHz LTE-A signal with a PAPR of 9.3 dB and a carrier-aggregated signal formed by four-carrier WCDMA and LTE-A signals with total bandwidth of 80 MHz are used. The DPD coefficients are extracted using the proposed baseband direct learning approach, and the process is iterated several times until convergence. A CRV model [43] is used as the DPD engine. Under-sampling factors of $N_u = 1, 2$ and 5, which represent complex ADC rates of 5x, 2.5x and 1x over the input signal bandwidth, are used with the direct learning. The output spectrum for the LTE-A and carrier aggregated signals are shown in Figs. 5.3 and 5.4.

Measurement results are summarized in Table 5.1, with both lower and upper ACLR reported. The proposed approach has comparable linearization capability for a wideband signal of up to 80 MHz even when the sampling rate is 1x the input signal bandwidth. By contrast, the best reported result in [84] is 2x.

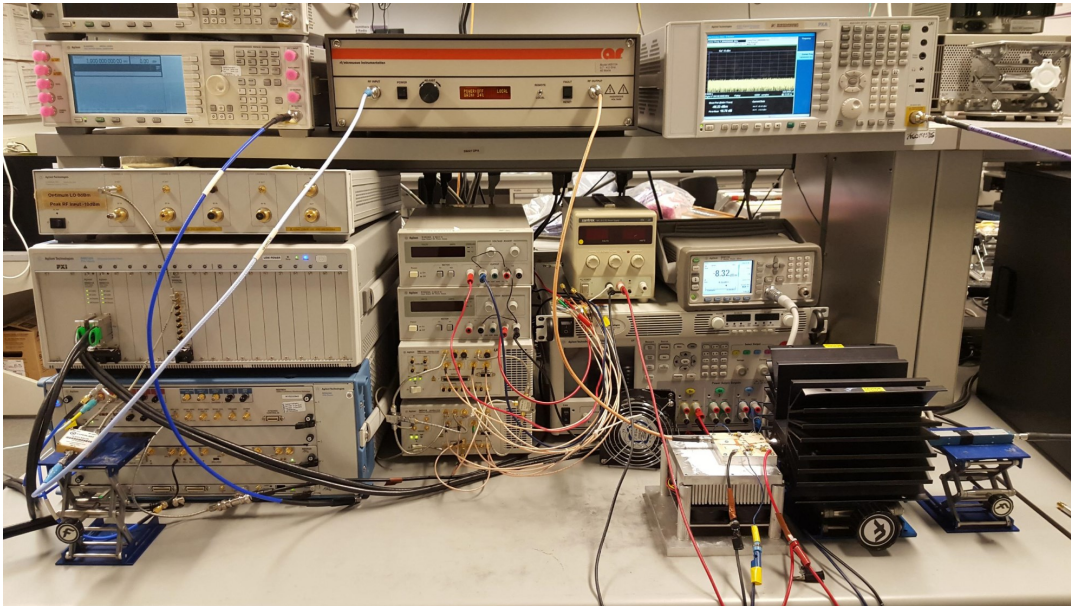
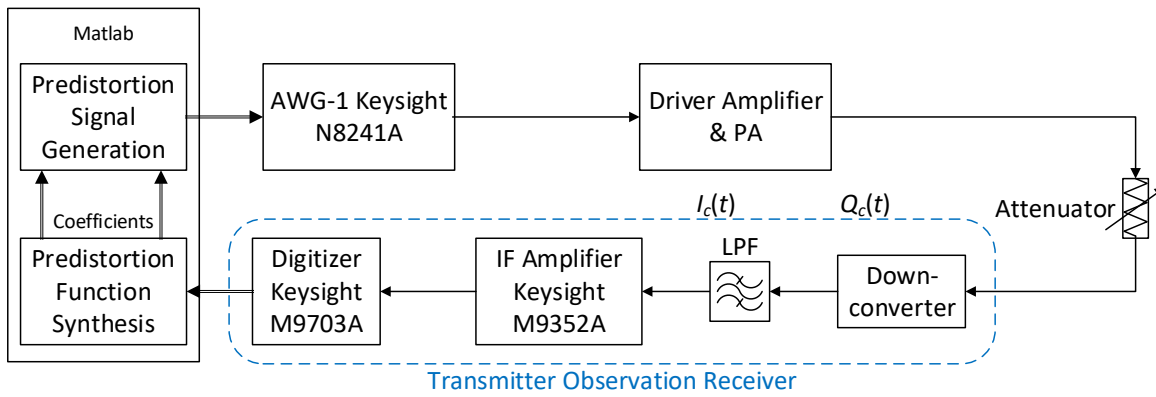


Figure 5.2: Measurement setup of the proposed baseband under-sampling scheme.

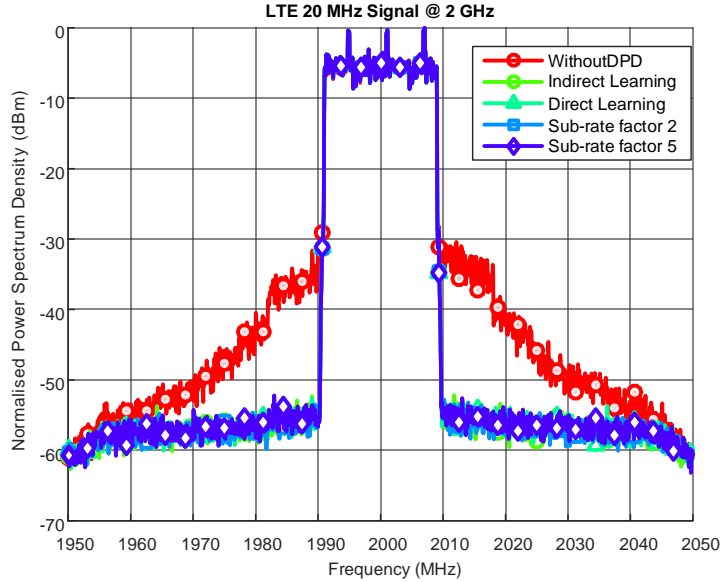


Figure 5.3: Spectrum of the PA output driven with the LTE-A signal.

5.6.2 Under-sampling using IF Scheme

The hardware optimized under-sampling training algorithm for the IF scheme is implemented together with the FPGA-based DPD system presented in Chapter 4. Together with the parallel processing DPD engine, it forms a complete solution of scalable DPD system capable of linearizing ultra-wideband signals.

Two test cases were used under different types of measurement setups. The first was a conducted measurement setup (i.e., measurements are performed via cables) using a silicon DPA as the DUT and OFDM signals with 400 MHz modulation bandwidth. The second was an OTA measurement setup with a 64-element beamforming array as the DUT and OFDM signals with 800 MHz modulation bandwidth.

The measurement setup is very similar to those presented in the measurement section in Chapter 4. The same baseband stage and up-converter highlighted in Fig. 4.8 are used to provide the mm-wave input to feed the array radio head placed inside the anechoic chamber. The DPD engine was the same and implemented using the programmable logic of the FPGA on a Xilinx Ultrascale+ MPSOC evaluation board (ZCU102), and operated at a core clock rate of 300 MHz, with scalable bandwidth up to 2.4 GHz.

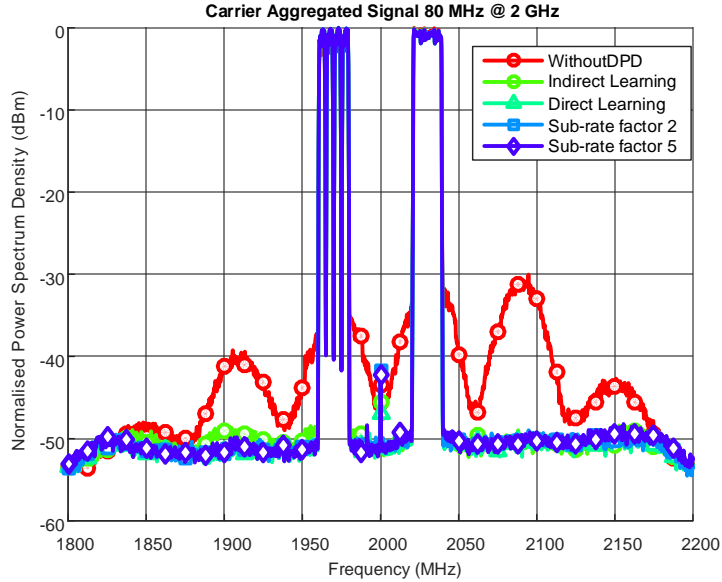


Figure 5.4: Spectrum of the PA output driven with the CA signal.

Fig. 5.5 shows the setup of the conducted measurements. The RF-DAC runs at a clock rate of 4.8 GSPS and uses a built-in digital-upconverter stage to translate the complex baseband signal to an IF frequency of 1.5 GHz. Then, the IF output from the RF-DAC is up-converted to 28 GHz using an image-rejection IQ up-converter followed by a driver amplifier before feeding to the DUT. The output of the DPA is down-converted to an IF frequency of 2.2 GHz and sampled by the RF-ADC (AD9208 from Analog Devices) at 200 MHz. This results in a maximum under-sampling factor of 24. These frequencies at the TOR portion were set to follow the constraints imposed by (5.46) to allow for a hardware-efficient implementation of an under-sampling TOR. It is worth noting that the under-sampling TOR can only be used to train the DPD coefficients and cannot be used to assess the performance metrics [e.g., EVM] at the PA output due to the aliasing of samples captured by the TOR. Therefore, a spectrum analyzer (Keysight N9040B) is used to capture the output spectrum and measure the ACPR and EVM.

Fig. 5.6 show the spectrum at the output of the DPA in the conducted measurements 400 MHz OFDM test signals formed by aggregating 4 component carriers (2-CC) of 100 MHz 64-QAM OFDM signals and trained with under-sampling TOR. Their ACPR and EVM are summarized in Table 5.2 and compared with results from Table 4.2. Ac-

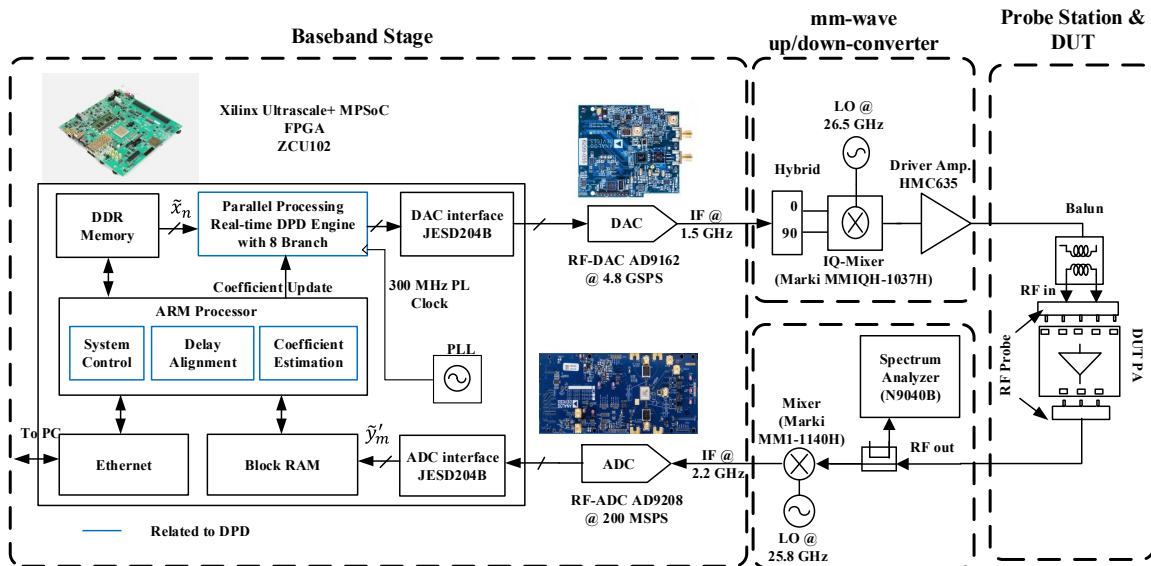


Figure 5.5: Measurement setup of proposed DPD system – linearizing a silicon DPA in a probe station.

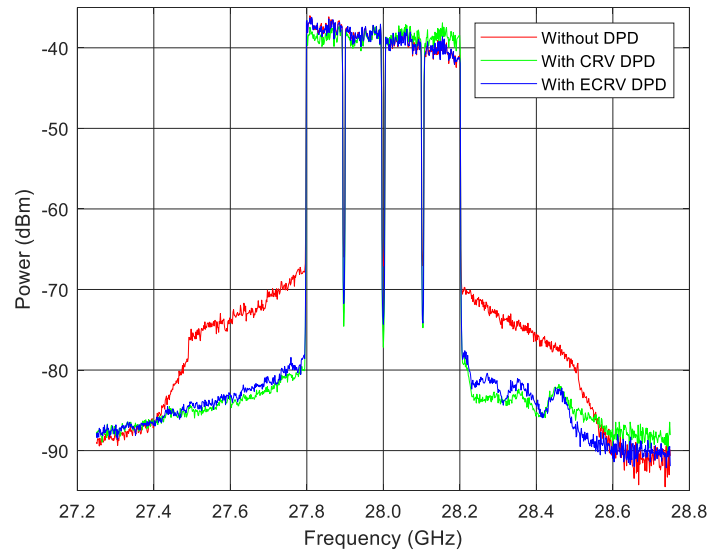


Figure 5.6: Output spectra of mm-wave silicon DPA in conducted measurement setup operating at 28 GHz using signals with 400 MHz bandwidths trained with under-sampling TOR.

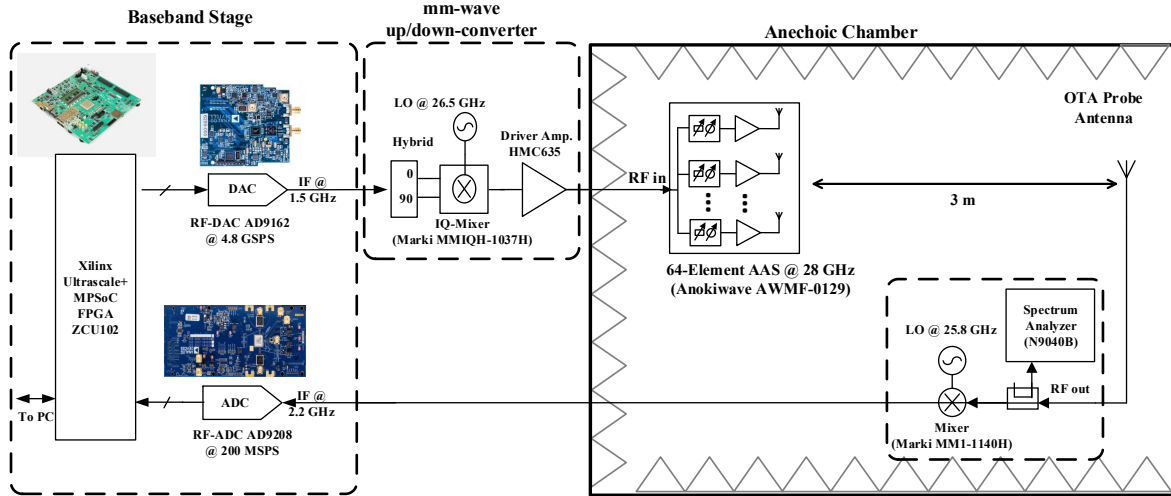


Figure 5.7: Measurement setup of proposed DPD system – linearizing a 64-element mm-wave beamforming array with OTA receiving antenna.

According to Table 5.2, similar extent of linearization performance is observed using either the CRV- or ECRV-based DPD engines trained with under-sampling compare to results obtained from full-rate training. The EVM reduced from 3.9% before DPD to 1.1% after DPD. The ACPR reduced from -32 dBc before DPD to around -45 dBc after DPD.

Similarly, an OTA measurement at 28 GHz is setup as shown in Fig. 5.7, with 800 MHz test signals. A spectrum analyzer was used to assess the performance metrics (ACPR and EVM) of the signal captured by the horn antenna.

The output spectra of the mm-wave 64-element array captured by the OTA horn antenna are shown in Fig. 4.13 for the 800 MHz test cases. The ACPR and EVM for both DPD engines are summarized in Table 5.2. Both DPD engines achieved EVM improvements from 4.1% to 1.2%, and ACPR improvements from -33 dBc to -40 dBc after DPD, compare to results obtained from full-rate training.

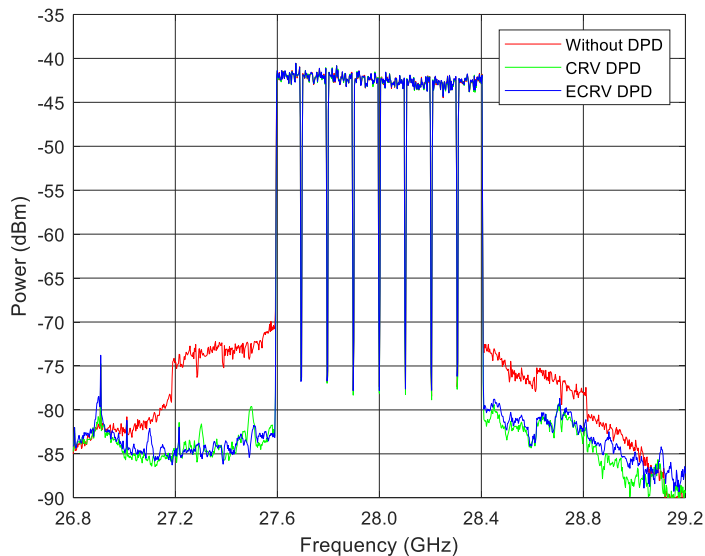


Figure 5.8: Output spectra of mm-wave 64-element beamforming array in the OTA measurement setup operating at 28 GHz using signals with 800 MHz bandwidth trained with under-sampling TOR.

Table 5.2: Measurement Results of IF Based Training with Under-sampled Feedback Signal

	DPA 400 MHz Signal		OTA 800 MHz Signal	
	ACPR	EVM	ACPR	EVM
No DPD	-32.6/-34.8	3.9%	-33.6/-36.5	4.1%
CRV Full-rate	-44.6/-46.1	1.1%	-40.7/-42.0	1.1%
CRV Under-sampled	-44.5/-46.1	1.1%	-40.7/-41.9	1.1%
ECRV Full-rate	-43.9/-45.5	1.1%	-40.2/-40.9	1.2%
ECRV Under-sampled	-43.9/-45.4	1.1%	-40.1/-40.9	1.2%

5.7 Conclusion

In this chapter, a new approach to identifying the coefficients of the DPD using under-sampled feedback signals was presented. Using a direct learning based architecture, the proposed approach was able to compute the coefficient updates of the predistortion engine using both the baseband and IF aliased samples of the PA output. Detailed algorithms for both baseband and IF schemes were presented, with hardware optimized implementation considerations discussed. In addition, a delay estimation and compensation algorithm were proposed to cancel the delays between the input and under-sampled output signals. The proposed approach was experimentally demonstrated to successfully synthesize the DPD function despite the presence of aliasing in the feedback signal due to significant under-sampling. Conducted on a 20 W GaN DPA driven by modulated signals with up to 80 MHz bandwidth, measurement results show comparable linearization performance for the proposed approach and a conventional full-rate indirect learning based DPD using the baseband scheme. The hardware optimized IF scheme was also implemented together with a parallel processing based DPD engine on an FPGA-base SOC, to provide a complete scalable DPD solution capable of linearizing ultra-wideband signals. It demonstrated over 2.4 GHz linearization bandwidth while the ADC was operating at a clock rate of 200 MHz. Its performance was demonstrated experimentally by linearizing a silicon DPA with 200 MHz and 400 MHz signals in conductive measurements, and a 64-element beam-forming array with 400 MHz and 800 MHz signals in OTA testing. Although the discussion in this chapter was presented in the context of a DPD system, the same principle can be applied in analog predistortion with only a little extra effort.

Chapter 6

Conclusion

Demand for high speed mobile data is driving the rapid development of mobile technology to use small-cell base stations as well as ultra-wideband signals in future sub-6 GHz and 5G mm-wave systems. The PA is the key component that dominates the linearity and efficiency of a transmitter. Conventional DPD systems used to linearize RF PAs are facing increasing challenges due to high power overhead for small-cell base stations, and the drastic increase in signal bandwidth that push beyond the capabilities of current digital circuits. This thesis aims to address such problems by introducing new approaches to predistortion systems to support future sub-6 GHz small-cell base stations and support wideband transmission in 5G mm-wave systems. It began by examining the linearity efficiency trade-off of RF PAs, and reviewing existing linearization solutions. There is renewed interest in analog/RF predistortion as a replacement for small-cell base stations, to take advantage of the low power nature of analog circuits, yet system level designs to reduce hardware complexity are still needed, particularly for analog implementation while maintaining good linearization capability. And the hardware implementation of the predistortion system once again becomes an important research focus as the required computation of the predistortion engine starts to exceed the limit provided by the digital circuits available. There are also significant developments in feedback and training methods, largely focused on breaking away from conventional Nyquist rate sampling of the PA output and enabling proper coefficient update with significantly reduced ADC sampling rates in the TOR.

To provide a viable solution for sub-6 GHz small-cell base stations, an DA-ARFPD system using the FIR-EMP model along with a linear SSAPI algorithm has been presented. An DA-ARFPD test bench, which incorporates major RF components, has been built to assess the validity of the proposed FIR-EMP scheme and the SSAPI algorithm. It has been demonstrated that the SSAPI algorithm can extract coefficients with excellent modelling

accuracy in a single iteration for cascaded blocks of the FIR-EMP. Measurement results have shown that the proposed FIR-EMP model using the SSAPI algorithm can successfully linearize multiple PAs driven with various wideband and carrier-aggregated signals of up to 80 MHz instantaneous bandwidth. Linearization performance comparable to a DDR-Volterra based DPD scheme indicates the viability of the proposed FIR-EMP model for implementing DA-ARFPD modules capable of mitigating the distortions exhibited by PAs driven by communication signals with up to 80 MHz modulation bandwidths. This confirms the potential of DA-ARFPD as a very promising candidate for the linearization of sub-6 GHz small-cell base station PAs, which would reduce the power overhead compared to using the popular DPD techniques.

Next, a hardware-efficient real-time DPD system with scalable linearization bandwidth for ultra-wideband sub-6 GHz and 5G mm-wave transmitters has been presented. Using a novel parallel-processing DPD engine architecture, multiple samples can be processed per clock cycle, overcoming the maximum linearization bandwidth limit imposed by the maximum clock frequency of the digital circuits. The proposed DPD system is able to efficiently linearize wideband signals with potentially unlimited bandwidths. Hardware design choices that optimize resource usage are presented using a sample designed on a commercial FPGA. The proposed DPD system achieved over 2.4 GHz linearization bandwidth using only a 300 MHz core clock for the digital circuits. Using a silicon DPA with 200 MHz and 400 MHz signals as the DUT for conductive measurements, and a 64-element beamforming array with 400 MHz and 800 MHz signals as the DUT for OTA testing, the scalability and linearization capability of the proposed system were demonstrated at 28 GHz. Experimental results show that the DPD engine using an envelope-based model such as ECRV can achieve similar performance to conventional DPD models like CRV for linearizing mm-wave DUTs transmitting wideband OFDM signals, while using less hardware and consuming less power.

Common to both solutions, a new approach to identify the coefficients of the DPD using an under-sampled feedback signal was presented. Using direct learning based architecture, the proposed approach is able to compute the coefficient updates of the predistortion engine using both the baseband and IF aliased samples of the PA output. Detailed algorithms for both baseband and IF schemes are presented, with hardware optimized implementation considerations discussed. In addition, a delay estimation and compensation algorithm has been proposed to cancel the delays between the input and under-sampled output signals. The proposed approach experimentally demonstrated its ability to successfully synthesize the DPD function despite the presence of aliasing in the feedback signal due to significant under-sampling. Conducted on a 20 W GaN DPA driven by modulated signals with up to 80 MHz bandwidth, measurement results show comparable linearization performance for

the proposed approach and the conventional full-rate indirect learning based DPD using a baseband scheme. The hardware optimized IF scheme was also implemented together with a parallel processing based DPD engine on FPGA-base SOC, to provide a complete scalable DPD system solution capable of linearizing ultra-wideband signals. It demonstrated over 2.4 GHz linearization bandwidth while the ADC was operating at a clock rate of 200 MHz.

6.1 Summary of Contributions and Publications

The goal of this thesis was to develop new linearization methods suitable for RF PAs used in future small-cell base stations and 5G mm-wave systems supporting wideband transmission. This goal was achieved through the following key contributions and publications:

- An DA-ARFPD system using the FIR-EMP model along with a linear SSAPI algorithm was presented which provides a low-power linearization solution for small-cell base stations.
 - H. Huang, A. Islam, J. Xia, P. Levine, and S. Boumaiza, "Linear filter assisted envelope memory polynomial for analog/radio frequency predistortion of power amplifiers," in *Proc. IEEE MTT-S IMS*, pp. 1-3, May. 2015.
 - H. Huang, J. Xia, A. Islam, E. Ng, P. M. Levine, and S. Boumaiza, "Digitally assisted analog/RF predistorter with a small-signal-assisted parameter identification algorithm," in *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 12, pp. 4297-4305, Dec. 2015
- A hardware-efficient real-time DPD system with scalable linearization bandwidth for wideband 5G mm-wave transmitters was developed. It removed the limits on the linearization bandwidth imposed by the digital circuitry maximum clock frequency.
 - H. Huang, J. Xia, and S. Boumaiza, "Parallel-processing-based digital predistortion architecture and FPGA implementation for wide-band 5G transmitters," in *2019 IEEE MTT-S International Microwave Conference on Hardware and Systems for 5G and Beyond (IMC-5G)*, Atlanta, GA, USA, 2019, pp. 1-3
 - H. Huang, J. Xia, and S. Boumaiza, "Novel parallel-processing-based hardware implementation of baseband digital predistorters for linearizing wideband 5G transmitters," in *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4066-4076, Oct. 2020

- A direct learning based under-sampling framework that enables the identification of the predistortion coefficients using under-sampled feedback signal was presented. It allowed the ADC to operate at a significantly lower clock rate and reduced the power and hardware resource usage of the TOR.
 - H. Huang, P. Mitran, and S. Boumaiza, "Digital predistortion function synthesis using undersampled feedback signal," in *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 10, pp. 855-857, Oct. 2016
 - H. Huang, J. Xia, and S. Boumaiza, "Novel parallel-processing-based hardware implementation of baseband digital predistorters for linearizing wideband 5G transmitters," in *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4066-4076, Oct. 2020

6.2 Future Work

The need for higher efficiency, better linearity and wider bandwidth in RF transmitters continues to grow and new challenges in PA linearization will emerge. Solutions presented in this work can be further extended in various ways to meet changing demands.

This work presented an DA-ARFPD system for sub-6 GHz small-cell base stations, yet development in pico- or femto-cells with even smaller output powers will require further reductions to the linearization system power overhead. Higher levels of integration with other components in the transmitter system, and realization of predistortion functions using nonlinear analog components, are both additional interesting topics to explore. The low power nature of DA-ARFPD systems can make them good candidates for MIMO systems that contain a large number of PAs. Adaptation of the ARFPD system to meet the power, space and system integration requirements of MIMO systems, as well as modelling and training the predistorter, present abundant research opportunities.

The parallel-processing DPD system presented in this thesis is designed for SISO DPD, suitable for phased-array or single-user single-frequency-band situations. It is natural to extend this work to MIMO systems (i.e., multi-user and multi-band scenarios) where dual- or multi-input DPD schemes are required. The linearity problem of hybrid or fully digital arrays under multi-user scenarios may require innovations in linearization approaches in terms of development of predistortion models and implementation approaches. Novel architectures will be required to achieve efficient use of hardware resources. The current architecture is based on commercial FPGA fabric using programmable logic cells and DSP units, and development in FPGA or SOC technology could provide other opportunities

to take advantage of the hardware. For example, FPGAs and ASIC designed for artificial intelligence provide a large array of memory-computing units and could be good candidates for novel DPD architectures. The design of the feedback path for predistortion also requires further attention. Conventional TORs using dedicated receiving antennas separated from the transmitting array are not practical and other solutions introduce new aliasing and interference into the feedback signal. New designs and algorithms need to be developed to address such issues.

Above all, the linear and efficient transmission of wideband signals should be addressed from a system-level approach that continues to increase the level of system integration. Linearization systems are the first step in this trend, and enable the relaxation of design constraints of RF PAs by addressing the linearity of the system at the transmitter level. With higher levels of integration in MIMO systems and other future developments in wireless communication, it is reasonable to expect that signal quality issues would be addressed at higher system levels. PA linearization systems such as predistortion can be combined with other features such as IQ imbalance to provide a more comprehensive solution. There is already work done that shows promising results, and further research in this direction could be an interesting topic.

References

- [1] Ericson, “Ericsson Mobility Report November 2019 excerpt,” https://www.ericsson.com/48f48f/assets/local/mobility-report/documents/2019/emr_handout-dec2019.pdf, Nov. 2019, [Online; accessed Dec-2019].
- [2] J. Hoydis, M. Kobayashi, and M. Debbah, “Green small-cell networks,” *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.
- [3] Maxim Integrated Product Inc., “RF predistortion (RFPD) vs. digital predistortion (DPD),” Sep. 2015.
- [4] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, “Beam-oriented digital predistortion for 5g massive MIMO hybrid beamforming transmitters,” *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018.
- [5] E. Ng, Y. Beltagy, G. Scarlato, A. B. Aayed, P. Mitran, and S. Boumaiza, “Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 11, pp. 4479–4492, Nov. 2019.
- [6] S. Cripps, *Advanced Techniques in RF Power Amplifier Design*. Artech House, 2002.
- [7] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*. Artech House, 2006.
- [8] J. Moon, J. Son, J. Lee, and B. Kim, “A multimode/multiband envelope tracking transmitter with broadband saturated amplifier,” *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3463–3473, Dec. 2011.

- [9] A. K. Kwan, M. Younes, R. Darraji, and F. M. Ghannouchi, “On track for efficiency: Concurrent multiband envelope-tracking power amplifiers,” *IEEE Microw. Mag.*, vol. 17, no. 5, pp. 46–59, May 2016.
- [10] H. Sarbishaei, “Concurrent multi-band envelope tracking power amplifiers for emerging wireless communications,” Ph.D. dissertation, University of Waterloo, 2014.
- [11] G. T. Watkins and K. Mimis, “How not to rely on Moore’s law alone: Low-complexity envelope-tracking amplifiers,” *IEEE Microw. Mag.*, vol. 19, no. 4, pp. 84–94, Jun. 2018.
- [12] W.-T. Tsai, C.-Y. Liou, Z.-A. Peng, and S.-G. Mao, “Wide-bandwidth and high-linearity envelope-tracking front-end module for LTE-a carrier aggregation applications,” *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4657–4668, Nov. 2017.
- [13] I. Kim, J. Moon, S. Jee, and B. Kim, “Optimized design of a highly efficient three-stage doherty PA using gate adaptation,” *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 10, pp. 2562–2574, Oct. 2010.
- [14] D. Y.-T. Wu and S. Boumaiza, “A modified doherty configuration for broadband amplification using symmetrical devices,” *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 10, pp. 3201–3213, Oct. 2012.
- [15] H. Golestaneh, F. A. Malekzadeh, and S. Boumaiza, “An extended-bandwidth three-way doherty power amplifier,” *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 9, pp. 3318–3328, Sep. 2013.
- [16] A. Jundi, H. Sarbishaei, and S. Boumaiza, “An 85-w multi-octave push–pull GaN HEMT power amplifier for high-efficiency communication applications at microwave frequencies,” *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 11, pp. 3691–3700, Nov. 2015.
- [17] G. Nikandish, R. B. Staszewski, and A. Zhu, “Breaking the bandwidth limit: A review of broadband Doherty power amplifier design for 5G,” *IEEE Microw. Mag.*, vol. 21, no. 4, pp. 57–75, Apr. 2020.
- [18] N. Sokal and A. Sokal, “Class E—a new class of high-efficiency tuned single-ended switching power amplifiers,” *IEEE Trans. Syst. Sci. Cybern.*, vol. 10, no. 3, pp. 168–176, Jun. 1975.

- [19] B. Berglund, J. Johansson, and T. Lejon, “High efficiency power amplifiers,” *Ericsson Review*, vol. 83, no. 3, pp. 92–96, 2006.
- [20] A. Birafane, M. El-Asmar, A. Kouki, M. Helou, and F. Ghannouchi, “Analyzing LINC systems,” *IEEE Microw. Mag.*, vol. 11, no. 5, pp. 59–71, Aug. 2010.
- [21] T. Barton, “Not just a phase: Outphasing power amplifiers,” *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 18–31, Feb. 2016.
- [22] M. Litchfield and T. Cappello, “The various angles of outphasing PAs: Competitiveness of outphasing in efficient linear PA applications,” *IEEE Microw. Mag.*, vol. 20, no. 4, pp. 135–145, Apr. 2019.
- [23] D. J. Perreault, “A new power combining and outphasing modulation system for high-efficiency power amplification,” *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 8, pp. 1713–1726, Aug. 2011.
- [24] C. M. Andersson, D. Gustafsson, J. C. Cahuana, R. Hellberg, and C. Fager, “A 1–3-GHz digitally controlled dual-RF input power-amplifier design based on a doherty-outphasing continuum analysis,” *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 10, pp. 3743–3752, Oct. 2013.
- [25] R. Staszewski, J. Wallberg, S. Rezeq, C.-M. Hung, O. Eliezer, S. Vemulapalli, C. Fernando, K. Maggio, R. Staszewski, N. Barton, M.-C. Lee, P. Cruise, M. Entezari, K. Muhamma, and D. Leipold, “All-digital PLL and transmitter for mobile phones,” *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2469–2482, Dec. 2005.
- [26] H. Wang, S. Kousai, K. Onizuka, and S. Hu, “The wireless workhorse: Mixed-signal power amplifiers leverage digital and analog techniques to enhance large-signal RF operations,” *IEEE Microw. Mag.*, vol. 16, no. 9, pp. 36–63, Oct. 2015.
- [27] J. S. Park, Y. Wang, S. Pellerano, C. Hull, and H. Wang, “A CMOS wideband current-mode digital polar power amplifier with built-in AM–PM distortion self-compensation,” *IEEE J. Solid-State Circuits*, vol. 53, no. 2, pp. 340–356, Feb. 2018.
- [28] S. Kousai and A. Hajimiri, “An octave-range, watt-level, fully-integrated CMOS switching power mixer array for linearization and back-off-efficiency improvement,” *IEEE J. Solid-State Circuits*, vol. 44, no. 12, pp. 3376–3392, Dec. 2009.
- [29] M. S. Alavi, R. B. Staszewski, L. C. N. de Vreede, and J. R. Long, “A wideband 2×13 -bit all-digital i/q RF-DAC,” *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 732–752, Apr. 2014.

- [30] S. Balasubramanian, S. Boumaiza, H. Sarbishaei, T. Quach, P. Orlando, J. Volakis, G. Creech, J. Wilson, and W. Khalil, "Ultimate transmission," *IEEE Microw. Mag.*, vol. 13, no. 1, pp. 64–82, Jan. 2012.
- [31] F. Wang, T.-W. Li, S. Hu, and H. Wang, "A super-resolution mixed-signal Doherty power amplifier for simultaneous linearity and efficiency enhancement," *IEEE J. Solid-State Circuits*, vol. 54, no. 12, pp. 3421–3436, 2019.
- [32] H. Sarbishaei, D. Y.-T. Wu, and S. Boumaiza, "Linearity of GaN HEMT RF power amplifiers - a circuit perspective," *2012 IEEE/MTT-S International Microwave Symposium Digest*, Jun. 2012.
- [33] H. Golestaneh, "Broadband doherty power amplifiers with enhanced linearity for emerging radio transmitters," Ph.D. dissertation, University of Waterloo, 2016.
- [34] J. C. Pedro, P. M. Cabral, T. R. Cunha, and P. M. Lavrador, "A multiple time-scale power amplifier behavioral model for linearity and efficiency calculations," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 606–615, Jan. 2013.
- [35] M. Hassan, L. E. Larson, V. W. Leung, and P. M. Asbeck, "Effect of envelope amplifier nonlinearities on the output spectrum of envelope tracking power amplifiers," in *2012 IEEE 12th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*. IEEE, Jan. 2012.
- [36] Y. Hu, "A novel power-scalable wideband power amplifier linearization technique," Ph.D. dissertation, University of Waterloo, 2019.
- [37] K. Hausmair, S. Gustafsson, C. Sanchez-Perez, P. N. Landin, U. Gustavsson, T. Eriksson, and C. Fager, "Prediction of nonlinear distortion in wideband active antenna arrays," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov. 2017.
- [38] H. Ku, M. McKinley, and J. Kenney, "Quantifying memory effects in RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 12, pp. 2843–2849, Dec. 2002.
- [39] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York:Wiley, 1980.
- [40] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

- [41] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, “A generalized memory polynomial model for digital predistortion of RF power amplifiers,” *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
- [42] A. Zhu, J. Pedro, and T. Brazil, “Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers,” *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.
- [43] F. Mkadem, M. Fares, S. Boumaiza, and J. Wood, “Complexity-reduced Volterra series model for power amplifier digital predistortion,” *Analog Integ. Circuits and Signal Process.*, vol. 79, no. 2, pp. 331–343, 2014.
- [44] B. Fehri and S. Boumaiza, “Baseband equivalent Volterra series for behavioral modeling and digital predistortion of power amplifiers driven with wideband carrier aggregated signals,” *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2594–2603, Nov. 2014.
- [45] F. Mkadem and S. Boumaiza, “Extended hammerstein behavioral model using artificial neural networks,” *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 4, pp. 745–751, Apr. 2009.
- [46] —, “Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion,” *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.
- [47] E. G. Lima, T. R. Cunha, and J. C. Pedro, “A physically meaningful neural network behavioral model for wireless transmitters exhibiting PM–AM/PM–PM distortions,” *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 12, pp. 3512–3521, Dec. 2011.
- [48] D. Hush and B. Horne, “Progress in supervised neural networks,” *IEEE Signal Process. Mag.*, vol. 10, no. 1, pp. 8–39, Jan. 1993.
- [49] F. Raab, P. Asbeck, S. Cripps, P. Kenington, Z. Popovic, N. Potheary, J. Sevic, and N. Sokal, “Power amplifiers and transmitters for RF and microwave,” *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 814–826, Mar. 2002.
- [50] P. B. Kenington, *High Linearity RF Amplifier Design*. Artech House, 2000.
- [51] N. Potheary, *Feedforward Linear Power Amplifiers*. Artech House, 1999.

- [52] H. Choi, Y. Jeong, C. D. Kim, and J. S. Kenney, "Efficiency enhancement of feedforward amplifiers by employing a negative group-delay circuit," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 5, pp. 1116–1125, May 2010.
- [53] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," *IEEE Trans. Circuits Syst.*, vol. 23, no. 5, pp. 285–291, May 1976.
- [54] J. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 374–382, 1990.
- [55] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, Nov. 2001.
- [56] A. Zhu, J. C. Pedro, and T. R. Cunha, "Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge," *IEEE Trans. Microw. Theory Techn.*, vol. 55, no. 5, pp. 813–821, May 2007.
- [57] B. Fehri and S. Boumaiza, "Baseband equivalent Volterra series for digital predistortion of dual-band power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 3, pp. 700–714, Mar. 2014.
- [58] E. Westesson and L. Sundstrom, "Low-power complex polynomial predistorter circuit in CMOS for RF power amplifier linearization," in *European Solid-State Circuits Conf. (ESSCIRC)*, Sep. 2001, pp. 486–489.
- [59] T. Rahkonen, O. Kursu, M. Riikola, J. Aikio, and T. Tuikkanen, "Performance of an integrated 2.1 GHz analog predistorter," in *Int. Workshop on Integr. Nonlinear Microw. and Milliw. Circuits*, Jan. 2006, pp. 34–37.
- [60] N. Mizusawa, S. Tsuda, T. Itagaki, and K. Takagi, "A polynomial-predistortion transmitter for WCDMA," in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2007, pp. 350–608.
- [61] A. Kidwai and B. Jalali, "Power amplifier predistortion linearization using a CMOS polynomial generator," in *IEEE Radio Freq. Integr. Circuits Symp.*, Jun. 2007, pp. 255–258.
- [62] S. Boumaiza, J. Li, M. Jaidane-Saidane, and F. Ghannouchi, "Adaptive digital/RF predistortion using a nonuniform LUT indexing function with built-in dependence

- on the amplifier nonlinearity,” *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 12, pp. 2670–2677, Dec. 2004.
- [63] W. Kim, K. Cho, S. Stapleton, and J. Kim, “Baseband derived RF digital predistortion,” *Electron. Lett.*, vol. 42, no. 8, pp. 468–470, Apr. 2006.
- [64] W. Woo, M. Miller, and J. Kenney, “A hybrid digital/RF envelope predistortion linearization system for power amplifiers,” *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 1, pp. 229–237, Jan. 2005.
- [65] R. N. Braithwaite, “Memory correction for a WCDMA amplifier using digital-controlled adaptive analog predistortion,” in *IEEE Radio and Wireless Symp.*, 2010, pp. 144–147.
- [66] F. Roger, “A 200mW 100MHz-to-4GHz 11th-order complex analog memory polynomial predistorter for wireless infrastructure RF amplifiers,” in *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, 2013, pp. 94–95.
- [67] A. Zhu, “Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers,” *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 737–744, Feb. 2015.
- [68] T. Liu, S. Boumaiza, and F. Ghannouchi, “Augmented Hammerstein predistorter for linearization of broad-band wireless transmitters,” *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 4, pp. 1340–1349, Jun. 2006.
- [69] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*. Artech House, 2014.
- [70] A. Katz, J. Wood, and D. Chokola, “The evolution of PA linearization: From classic feedforward and feedback through analog and digital predistortion,” *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 32–40, Feb. 2016.
- [71] Y.-J. Liu, J. Zhou, W. Chen, and B.-H. Zhou, “A robust augmented complexity-reduced generalized memory polynomial for wideband RF power amplifiers,” *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2389–2401, May 2014.
- [72] F. Mkadem, A. Islam, and S. Boumaiza, “Multi-band complexity-reduced generalized-memory-polynomial power-amplifier digital predistortion,” *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 6, pp. 1763–1774, Jun. 2016.

- [73] O. Hammi, F. Ghannouchi, and B. Vassilakis, “A compact envelope-memory polynomial for RF transmitters modeling with application to baseband and RF-digital predistortion,” *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 5, pp. 359–36, May 2008.
- [74] S. Bassam, M. Helaoui, and F. Ghannouchi, “Crossover digital predistorter for the compensation of crosstalk and nonlinearity in MIMO transmitters,” *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1119–1128, May 2009.
- [75] S. Amin, P. N. Landin, P. Handel, and D. Ronnow, “Behavioral modeling and linearization of crosstalk and memory effects in RF MIMO transmitters,” *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 810–823, Apr. 2014.
- [76] F. M. Barradas, P. M. Tome, J. M. Gomes, T. R. Cunha, P. M. Cabral, and J. C. Pedro, “Power, linearity, and efficiency prediction for MIMO arrays with antenna coupling,” *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 5284–5297, Dec. 2017.
- [77] K. Hausmair, P. N. Landin, U. Gustavsson, C. Fager, and T. Eriksson, “Digital predistortion for multi-antenna transmitters affected by antenna crosstalk,” *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1524–1535, Mar. 2018.
- [78] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, “Beam-oriented digital predistortion for 5g massive MIMO hybrid beamforming transmitters,” *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018.
- [79] E. Ng, Y. Beltagy, G. Scarlato, A. B. Ayed, P. Mitran, and S. Boumaiza, “Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 11, pp. 4479–4492, Nov. 2019.
- [80] D. Zhou and V. E. DeBrunner, “Novel adaptive nonlinear predistorters based on the direct learning algorithm,” *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 120–133, Jan. 2007.
- [81] Y. Liu, J. J. Yan, H.-T. Dabag, and P. M. Asbeck, “Novel technique for wideband digital predistortion of power amplifiers with an under-sampling ADC,” *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2604–2617, Nov. 2014.

- [82] R. N. Braithwaite, “Closed-loop digital predistortion (DPD) using an observation path with limited bandwidth,” *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 726–736, Feb. 2015.
- [83] C. Yu, L. Guan, E. Zhu, and A. Zhu, “Band-limited volterra series-based digital predistortion for wideband RF power amplifiers,” *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 12, pp. 4198–4208, Dec. 2012.
- [84] L. Ding, F. Mujica, and Z. Yang, “Digital predistortion using direct learning with reduced bandwidth feedback,” in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2013, pp. 1–3.
- [85] A. Prata, D. C. Ribeiro, P. M. Cruz, A. S. R. Oliveira, and N. B. Carvalho, “RF subsampling feedback loop technique for concurrent dual-band PA linearization,” *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 12, pp. 4174–4182, Dec. 2016.
- [86] H. Huang, P. Mitran, and S. Boumaiza, “Digital predistortion function synthesis using undersampled feedback signal,” *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 10, pp. 855–857, Oct. 2016.
- [87] Z. Wang, L. Guan, and R. Farrell, “Undersampling observation-based compact digital predistortion for single-chain multiband and wideband direct-to-RF transmitter,” *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 5274–5283, Dec. 2017.
- [88] Y. Beltagy, P. Mitran, and S. Boumaiza, “Direct learning algorithm for digital predistortion training using sub-nyquist intermediate frequency feedback signal,” *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 1, pp. 267–277, Jan. 2019.
- [89] N. Guan, N. Wu, and H. Wang, “Digital predistortion of wideband power amplifier with single undersampling ADC,” *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 11, pp. 1016–1018, Nov. 2017.
- [90] J. Chani-Cahuana, M. Ozen, C. Fager, and T. Eriksson, “Digital predistortion parameter identification for RF power amplifiers using real-valued output data,” *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 10, pp. 1227–1231, Oct. 2017.
- [91] Q. Luo, X.-W. Zhu, C. Yu, and W. Hong, “Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk,” *IEEE Trans. Microw. Theory Techn.*, pp. 1–15, 2019.

- [92] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Trans. Microw. Theory Techn.*, pp. 1–14, 2019.
- [93] L. Guan and A. Zhu, "Low-cost FPGA implementation of volterra series-based digital predistorter for RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 866–872, Apr. 2010.
- [94] Y. Li, X. Wang, and A. Zhu, "Sampling rate reduction for digital predistortion of broadband RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 3, pp. 1054–1064, Mar. 2020.
- [95] C. Cheang, P. Mak, and R. P. Martins, "A hardware-efficient feedback polynomial topology for DPD linearization of power amplifiers: Theory and FPGA validation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 9, pp. 2889–2902, Sep. 2018.
- [96] Q. A. Pham, D. Lopez-Bueno, T. Wang, G. Montoro, and P. L. Gilabert, "Partial least squares identification of multi look-up table digital predistorters for concurrent dual-band envelope tracking power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5143–5150, Dec. 2018.
- [97] Y. Li, W. Cao, and A. Zhu, "Instantaneous sample indexed magnitude-selective affine function-based behavioral model for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, pp. 1–11, 2018.
- [98] H. Huang, A. Islam, J. Xia, P. Levine, and S. Boumaiza, "Linear filter assisted envelope memory polynomial for analog/radio frequency predistortion of power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, May 2015, pp. 1–3.
- [99] C. C. Cadenas, J. R. Tosina, M. J. M. Ayora, and J. M. Cruzado, "A new approach to pruning Volterra models for power amplifiers," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2113–2120, 2010.
- [100] E. G. Lima, T. R. Cunha, and J. C. Pedro, "PM-AM/PM-PM distortions in wireless transmitter behavioral modeling," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2011, pp. 1–4.
- [101] J. Pedro and S. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

- [102] M. N. A. Abadi, “Extended bandwidth Doherty power amplifier for carrier aggregated signals,” Master’s thesis, University of Waterloo, 2014.
- [103] H. S. A. Jundi and S. Boumaiza, “An 85-W multi-octave push-pull GaN HEMT power amplifier for high efficiency communication applications at microwave frequencies,” *IEEE Trans. Microw. Theory Techn.*, vol. PP, no. 99, pp. 1–10, Sep. 2015.
- [104] *UltraScale Architecture DSP Slice User Guide*, UG579(v1.9) ed., Xilinx Inc., Sep. 2019.
- [105] Y. Beltagy, A. Chung, P. Mitran, and S. Boumaiza, “On the calibration of the feedback receiver using reduced sampling rate and its application to digital predistortion of 5G power amplifiers,” in *IEEE MTT-S Int. Microw. Symp. Dig.* IEEE, Jun. 2017.