Linearizing Radio Frequency Power Amplifiers Using an Analog Predistortion Technique

by

Anik Islam

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2015

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.

Anik Islam

Abstract

As critical elements of the physical infrastructure that enables ubiquitous wireless connectivity, radio frequency power amplifiers (RFPAs) are constantly pushed to the limits of linear but efficient operation. Digital predistortion, as a means of circumventing the limitations of this inherent linearity – efficiency trade-off, has been a subject of prolific research for well over a decade. However, to support the unrestrained growth of broadband mobile traffic, wireless networks are expected to rely increasingly on heterogeneously-sized small cells which necessitate new predistortion solutions operating at a fraction of the power consumed by digital predistortion approaches.

This thesis pertains to an emerging area of research involving analog predistortion (APD) – a promising, low-power alternative to digital predistortion (DPD) for future wireless networks. Specifically, it proposes a mathematical function that can be used by the predistorter to linearize RFPAs. As a preliminary step, the challenges of transitioning from DPD to APD are identified and used to formulate the constraints that APD imposes on the predistorter function. Following an assessment of the mathematical functions commonly used for DPD, and an analysis of the physical mechanisms of RFPA distortion, a new candidate function is proposed. This function is both compatible with and feasible for an APD implementation, and offers competitive performance against more complex predistorter functions (that can only be implemented in DPD).

The proposed predistorter function and its associated coefficient identification procedure are experimentally validated by using them to linearize an RFPA stimulated with single-band carrier aggregated signals of progressively wider bandwidths. The solution is then extended to the case of dual-band transmission, and subsequently validated on an RFPA as well. The proposed function is a cascade of a finite impulse response filter and an envelope memory polynomial and has the potential to deliver far better linearization results than what has been demonstrated to date in the APD literature.

Acknowledgements

I would like to thank Dr. Boumaiza for providing me with opportunities to learn and grow both professionally and personally. To Peter, Hai, Bilel, Hassan, and Farouk – I would not be here without your guidance and constructive criticism. To Ammu, Anna khala and Mahbub uncle – thanks for taking care of me; I hope to return the favour someday.

Last, but not least, to Brian and Sarah – thanks for putting up with me; to you two, I dedicate my thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

ACPR - L/U	Adjacent Channel Power Ratio – Lower/Upper
ADC	Analog-to-Digital Converter
AM/AM	Amplitude-to-Amplitude Modulation
AM/PM	Amplitude-to-Phase Modulation
APD	Analog Predistortion/Predistorter
AWG	Arbitrary Waveform Generator
BBE	Baseband Equivalent
DAC	Digital-to-Analog Converter
DC	Direct Current
DDR	Dynamic Deviation Reduction
DPD	Digital Predistortion/Predistorter
DSP	Digital Signal Processor
EMP	Envelope Memory Polynomial
EVM	Error Vector Magnitude
FIR	Finite Impulse Response
FIR||EMP	Envelope Memory Polynomial with parallel Finite Impulse Response Filter
FIR-APD	Analog Predistorter with cascaded Finite Impulse Response Filter
FIR-EMP	Envelope Memory Polynomial with cascaded Finite Impulse Response Filter
GaAs	Gallium Arsenide
GaN	Gallium Nitride
LSE	Least Squares Estimation
LTE-A	Long Term Evolution – Advanced
NMSE	Normalized Mean Squared Error
PAPR	Peak-to-Average-Power Ratio
RFPA	Radio Frequency Power Amplifier
RLS	Recursive Least Squares
TOR	Transmitter Observation Receiver
WCDMA	Wideband Code Division Multiple Access

Chapter 1
Introduction

1.1 Background

For designers of radio frequency power amplifiers (RFPAs), the inherent trade-off between efficiency and linearity has always been a challenge. With increasing pressure on the Information and Communication Technology (ICT) sector to reduce its carbon footprint, new ways are being sought to improve the average efficiencies of RFPAs, which consume the most power in RF front ends. Efficiency improvement techniques that are gaining traction, such as Envelope Tracking [1] and Doherty [2], employ transistor topologies and modes of operation that compromise the linearity of RF front ends, leading to unwanted spectral emissions and degraded signal quality at the transmitter. Concurrently, the need to support higher data rates in an increasingly connected society has resulted in communication signals experiencing a widening of modulation bandwidth and an increase of peak to average power ratio (PAPR), which aggravates nonlinearity and memory effects in RFPAs. These trends have necessitated explicit measures to improve the linearity of RFPAs by counteracting their nonlinearity and memory effects. Based on the approach, the measures can be broadly categorized as (i) feedback linearization, (ii) feed-forward linearization, and (iii) predistortion. Each of these is briefly reviewed below.

Despite its conceptual simplicity and highly adaptive nature, feedback linearization [3] is prone to instability, and is therefore limited to narrow bandwidths and low frequencies of operation. Feed-forward linearization [4] experiences no such direct bandwidth limitation but requires meticulous design and incurs considerable power overhead, making it an unattractive solution. Several approaches exist under the broad classification of predistortion, all of which involve some means of predistorting the RFPA input signal such that its distortions negate those generated by the amplifier to create a distortion-free output. One approach is to cascade nonlinear semiconductor devices (e.g., Schottky diodes [5] or FET transistors [6]) before the RFPA and judiciously configure them to predistort the RF signal. While this approach is viable for wideband signals and consumes less power relative to other solutions, it generally yields limited linearization success and is not adequate for solid-state RFPAs driven with wideband modulated, high PAPR signals. For these applications, digital predistortion (DPD) has become the approach of choice. DPD involves the use of digital signal processing (DSP) to deliberately introduce distortions in a baseband signal, which is then up-converted and fed into the RFPA. The predistortion is generated by a mathematical ‘inverse’ model of the PA, identified using a suitable learning algorithm. When the identification is done correctly, cascading this inverse model before the RFPA results in a distortion-free output.

Due to advancements in the processing power, programmability and size of digital circuitry, and the ease with which it can be integrated into the existing baseband processing of front ends, DPD has become a popular solution for linearizing RFPAs in wireless network base-stations. It has received considerable attention in the literature as well, as evidenced by the

myriad of papers discussing learning algorithms [7, 8], mathematical formulations ranging from the simple memory polynomials [9, 10] to the complex and powerful family of Volterra series [11, 12], and even meta-analyses contrasting different DPD solutions [13, 14].

1.2 Emerging Trends

As wireless networks evolve from 3G to 4G to support the needs of an increasingly data-hungry digital society, the status of DPD as the predominant linearization technique is being challenged. RF front ends must adapt to emerging trends – such as the utilization of ultra-high frequencies, concurrent transmission in multiple frequency bands, and deployment of small cells in heterogeneous networks – that are essential to realizing low-latency, high throughput, and ubiquitous wireless connectivity. As will be discussed in subsequent chapters, DPD will not continue to be viable going forward, due to reasons of excessive power consumption of its digital circuitry. The search for low-power DPD-alternatives has already prompted research efforts in which the minimization or complete removal of digital circuitry is a common theme.

Among the endeavors to realize digital-free predistortion, [15] is an attempt to partially reduce power consumption by shifting the burden of predistortion from digital baseband to the RF domain. While not without merit, this approach is ill-equipped to linearize RFPAs with memory effects. In [16], digital circuitry in the core of the predistorter is replaced entirely with analog circuit blocks, and the capability to address limited memory effects is introduced. Despite being the first prototype of a truly analog predistorter, the assumptions used in its design make it unsuitable for linearizing RFPAs in modern communication scenarios where memory effects can be significant. More recently, the first instance of packaged, fully integrated, analog predistortion (APD) circuit was reported in [17], with an unprecedented low power consumption of 200 mW. While full details of the solution have not been disclosed, evidence suggests that it employs an analog circuit realization of an envelope memory polynomial (EMP) [10]. The objective of this thesis is to propose a different mathematical formulation than EMP, which is demonstrably inadequate for the linearization of RFPAs transmitting wideband modulated signals. The architecture of the analog predistorter is scrutinized to reveal its limitations, which are taken into account in developing the proposed formulation. The efficacy of the proposed solution in linearizing a physical RFPA under wideband signal stimuli is demonstrated, and it is extended for concurrent dual-band transmission scenarios, followed by similar demonstrations. Methods and challenges of predistorter model identification are addressed as well.

1.3 Thesis Organization

The rest of this thesis is organized as follows. Chapter 2 provides the required background to appreciate the causes and effects of nonlinearity and memory effects in RFPAs, and describes in detail how DPD works to counteract them. The challenges faced in trying to adapt DPD systems to evolving RF front ends are highlighted, and the motivation and approach to seeking alternative solutions, with the aim of reducing power consumption of the overall predistorter, are discussed. APD is considered as a viable alternative.

Chapter 3 begins with an analysis of the APD architecture and its compatibility with pre-existing DPD schemes. Only the envelope memory polynomial is found to satisfy the constraints imposed by the APD architecture, but it is found to be inadequate for predistorting wideband communication signals. The performance of the EMP is considerably improved by cascading it after a finite impulse response (FIR) filter, but this complicates the identification of the coefficients. Two solutions for coefficient identification are presented, with trade-offs between computational cost and accuracy. The performance of the proposed FIR-EMP formulation is validated with measurement results on a physical RFPA, and compared against that of other standard DPD formulations.

In Chapter 4, the single-band FIR-EMP solution and the associated coefficient identification solutions are extended to the dual-band case. Changes to the theoretical formulation and to the experimental test setup are described. The efficacy of the proposed dual-band solution, compared to competing dual-band DPD formulations, is then demonstrated on the same RFPA as used in the single-band case.

Chapter 5 concludes the thesis with a discussion of the benefits and caveats of the proposed solution, and a discussion of future work to enhance and extend the solution for more challenging scenarios.

Chapter 2
Background

This chapter is organized as follows. Section 2.1 discusses the causes and effects of non-ideal RFPA behaviors (which are often aggravated in the pursuit of efficient operation) and introduces the figures of merit commonly used to assess their severity, and impact on signal integrity. Section 2.2 provides a review of digital predistortion — a popular method of correcting the impact of RFPA nonlinearity and memory effects — and discusses learning architectures, predistorter formulations, and coefficient identification algorithms associated with DPD. Section 2.3 highlights the issue of excessive power consumption in the conventional DPD architecture, and underscores the need for an alternative, low-power predistortion scheme which is described in Chapter 3.

2.1 Nonlinearity and Memory Effects in RFPAs

The pursuit of efficiency in RFPAs bears an antagonistic relationship with the pursuit of linearity. This is perhaps best illustrated through Figures 1 and 2, which show how efficiency and nonlinearity vary as a function of the RFPA class of operation (indicated by the conduction angle). Efficiency can be improved by increasing the conduction angle (Figure 1), but not without incurring penalties in linearity, as evidenced by the growing amplitude of harmonic components (Figure 2).

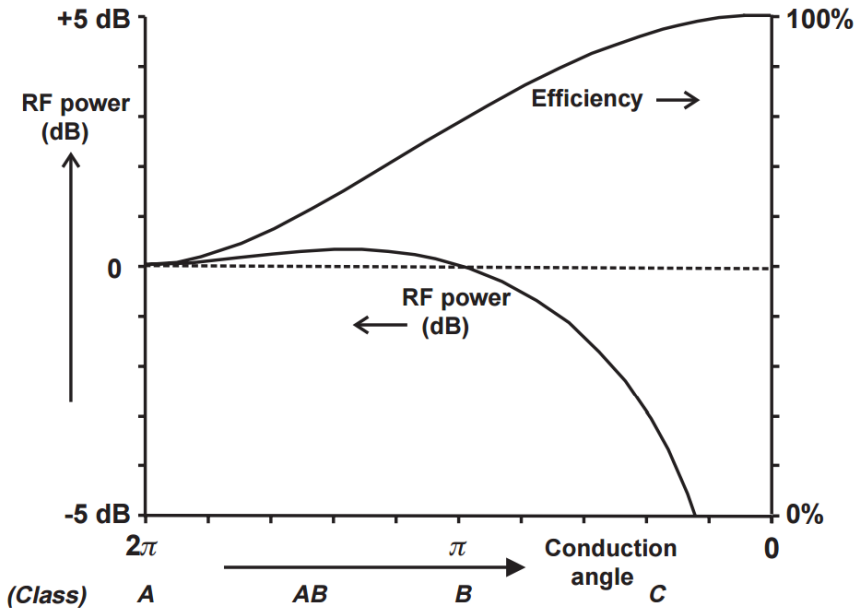

Figure 1: Efficiency as a function of the class of operation [18]

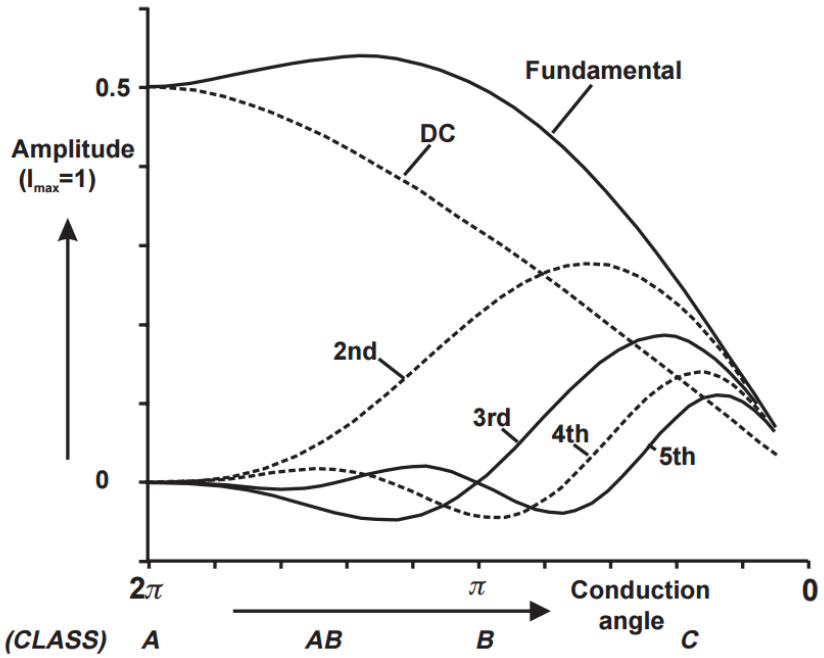

Figure 2: Nonlinearity as a function of the class of operation [18]

To efficiently transmit modulated signals with high PAPR, designers must ensure that RFPAs are efficient not only at peak power, but also at back-off levels closer to the average power of the signal. Techniques such as load modulation (Doherty) and drain modulation (Envelope Tracking) are commonly used to achieve this, but at the cost of linearity. The Doherty technique employs ‘class AB’– main and ‘class C’– auxiliary amplifiers, while the ET technique employs non-constant drain voltage; both approaches contribute to nonlinear behavior that results in new frequency components appearing at the RFPA output. Frequency components generated by harmonic distortion appear at multiples of the carrier frequencies, while those generated by intermodulation distortion (as a result of interactions between multiple input frequency components) appear around the carriers and close to the intended frequency of operation, as shown in Figure 3. Without countermeasures, these distortion products can interfere with out-of-band communication, and degrade communication in-band as well.

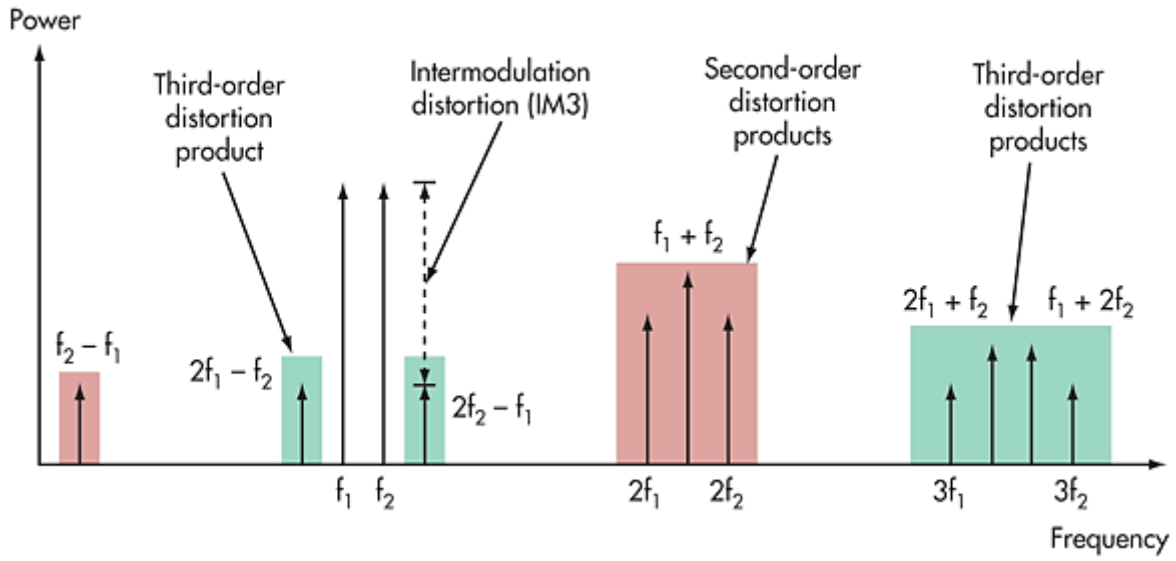

Figure 3: Distortion products created by a nonlinear RFPA [19]

Even after purely static nonlinearities are compensated, RFPA's may still exhibit distortion products that vary as a function of carrier spacing and modulation bandwidth – these are caused by memory effects. While not introducing new distortion products, memory effects introduce dynamic behavior in the amplitude and phase of existing distortion products. Based on the physical phenomena that cause them, memory effects can be categorized as:

- Electrical memory effects [20]**,** which are caused by interactions between the transistor's distortion products and the surrounding matching and bias networks (which cannot always be designed to have flat amplitude and phase responses over frequency).
- Thermal memory effects [20]**,** which are attributable to the dynamic self-heating of the RFPA and the change in electrical properties (such as gain) that occur as a result. While there is some debate as to whether the time constant of thermal effects is short enough to result in transient behavior during communication, back-of-the-envelope calculations suggest they can be of the order of microseconds, which is well within the timescale of modulated signal variation [21].
- Semiconductor trapping effects, which are the result of electron-bandgap interactions within the active element of the RFPA [22]. Their contribution to memory effects is less significant compared to electrical and thermal effects in most RFPAs. They are more prevalent in exotic semiconductors such as GaAs.

Another categorization of memory effects distinguishes between linear and nonlinear memory, based on the electrical circuit phenomenon responsible. Linear memory effects are caused by the frequency dispersive nature of matching networks around the transistor, and would distort the communication signal even in the absence of transistor nonlinearity. Nonlinear memory

effects, however, exist as a consequence of the transistor's nonlinear distortion products. When signals with multiple frequency components are subjected to the transistor's transfer characteristic, they give rise to intermodulation products, some of which appear at the difference frequencies of the various tones. In the case of signals with closely spaced tones (i.e. for the majority of modulated communication signals), some of these difference frequencies are located in the baseband (i.e they are relatively close to the direct current, or DC, as opposed to being in the passband, where the RF carrier is located). These baseband distortion currents cannot be filtered out by the bias network of the transistor which supplies it with DC power, and they interact with the drain impedance to cause low frequency fluctuations of the drain voltage. This 'accidental' drain modulation changes the nonlinear behavior of the transistor from one moment to the next, giving rise to nonlinear memory. The distinction between linear and nonlinear memory is recognized in the literature [23, 24], and becomes particularly important later in this thesis because of its implications on the mathematical inverse model of the RFPA that is used for predistortion.

A useful way to visualize the distortions introduced by the power amplifier (PA) is to plot an amplitude-to-amplitude distortion (AM/AM) characteristic, which shows how the power of the output signal P_{out} varies as a function of the input signal's power P_{in}. For purposes of characterization, the signals used to generate the AM/AM plot are typically normalized to an average power level of 0 dBm, and the ratio of the powers is plotted on the y-axis to show distortions in gain relative to a baseline of 0 dB. A similar visualization can be done for the amplitude-to-phase distortion (AM/PM) characteristic, which shows how the phase of the output signal phase $Phase_{out}$ varies as a function of the input signal's power. For purposes of characterization, the signals used to generate the AM/PM plot are typically delay-adjusted to have 0^{o} average phase difference, so that distortions in the phase shift are displayed relative to a baseline of 0^{o}.

Figure 4 illustrates the AM/AM and AM/PM of a Gallium Nitride (GaN) Doherty PA, when characterized with a long term evolution (LTE) signal of 40 MHz modulation bandwidth and 10.4 dB PAPR. It should be noted that these plots are not a complete representation of the power amplifier's transfer characteristic, since they are only valid for (i) the range of power levels over which the RFPA was driven, (ii) the particular modulated signal used to excite it, and (iii) its bias conditions at the time. Nonetheless, the curvature in the AM/AM and AM/PM (where ideally one would expect flat lines for a constant amplitude and phase characteristic across the range of input power) clearly illustrates nonlinear behavior, while the dispersion in plot data suggests that amplitude and phase distortions are not only a function of the input signal magnitude at any given time, but past times as well (memory effects).

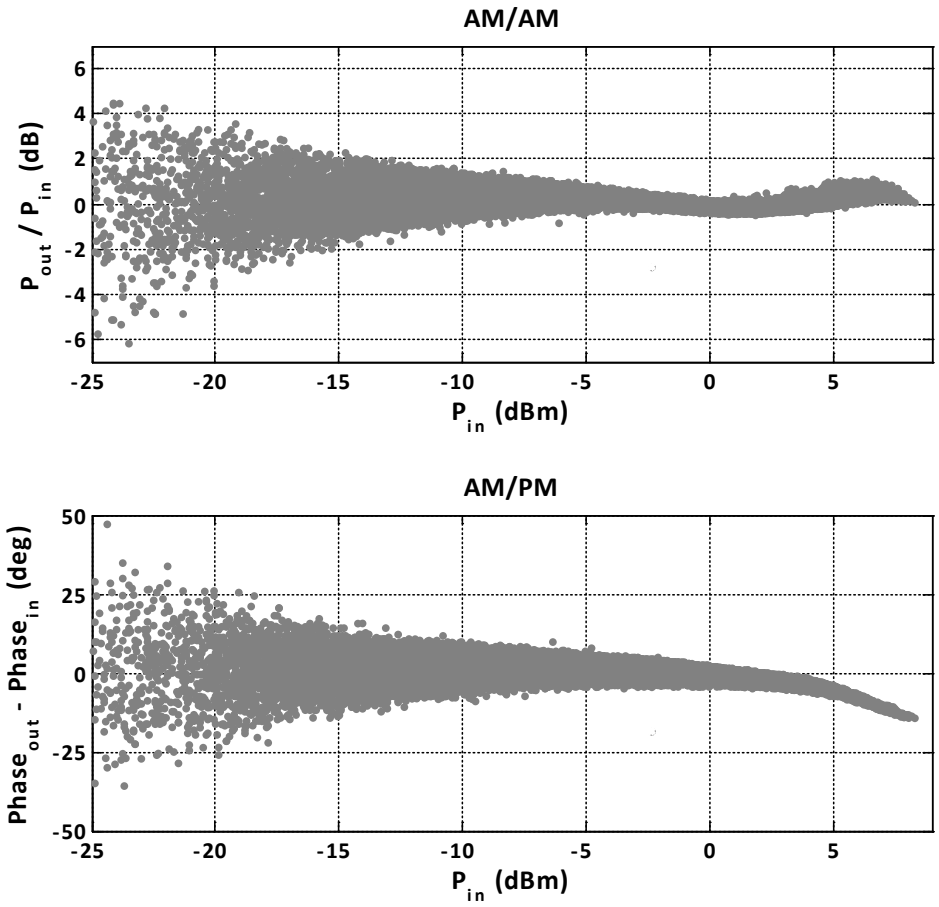

Figure 4: AM/AM (top) and AM/PM (bottom) of a Doherty PA driven by an LTE signal

As mentioned, the nonlinear behavior of the RFPA gives rise to both harmonic distortion products and intermodulation products. While the former can be removed using filters designed to reject the appropriate frequencies, the latter appear adjacent to the communication band of interest and must be addressed by linearization. A figure of merit to assess the severity of spectral distortion is the adjacent channel power ratio (ACPR, Eq. 2.1) which measures the logarithmic ratio of total power in a specified adjacent band P_{adj} and the total power within the communication band P_{in} (specified in dB). Reduction of out-of-band distortion products leads to a more negative ACPR, which is desirable. Figure 5 shows a power spectral density plot of the output of a Doherty PA with and without linearization, which clearly demonstrates the reduction in ACPR that can be realized with predistortion. While the ACPR is an indicator of out-of-band distortions, it yields no insight regarding in-band distortions, which affect the integrity of the data being transmitted. The error vector magnitude (EVM, Eq. 2.2) is used to quantify the latter, and is a measure of the magnitude of error introduced in the baseband IQ data $\sqrt{P_{err}}$, relative to the magnitude of the error-free data $\sqrt{P_{ref}}$ (specified in %). Reduction of in-band distortion leads to a smaller EVM, which is desirable. Both EVM and ACPR

reduction are important measures of success for linearization schemes, as a reduction of one of them does not necessarily translate to a reduction in the other.

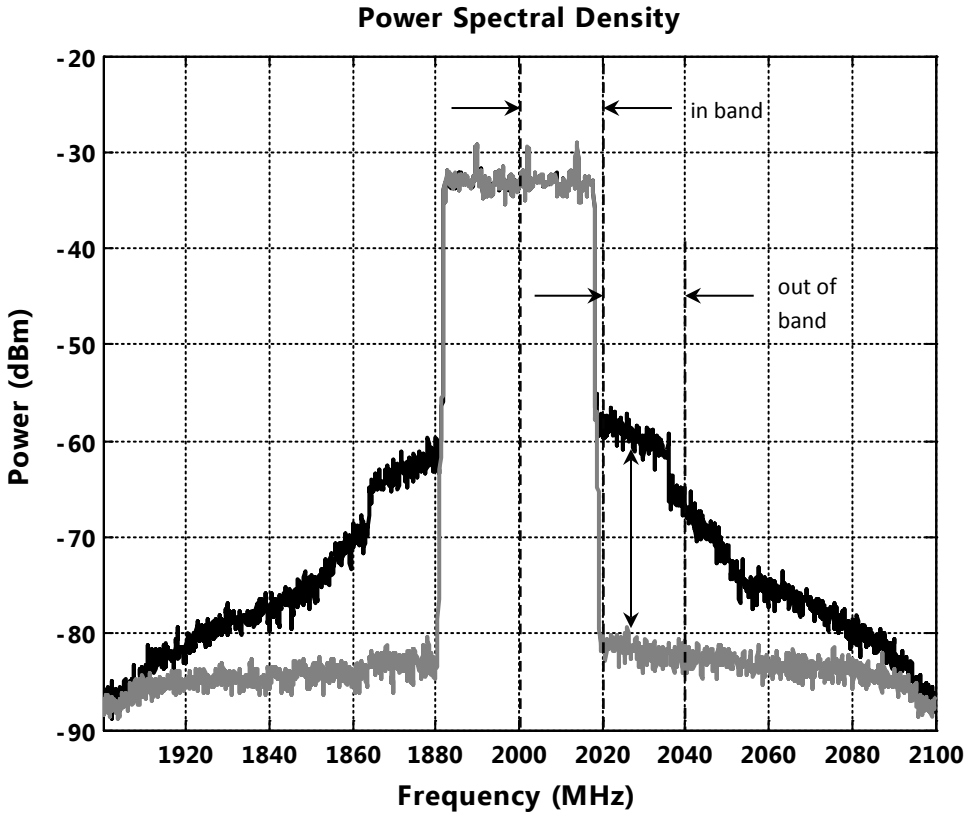

Figure 5: Power spectral density of Doherty PA output, showing out-of-band distortions with (grey) and without (black) linearization

$$ACPR = 10\log_{10}\left(\frac{P_{adj}}{P_{in}}\right) \quad (2.1) \qquad EVM = \sqrt{\frac{P_{err}}{P_{ref}}} \times 100\% \quad (2.2)$$

2.2 Linearization using Digital Predistortion

Regardless of the mathematical model and learning algorithm used, all digital predistortion schemes share certain principles of operation and a common architecture (Figure 6). A digital signal processor (the training engine) runs an estimation algorithm that calculates the coefficients of the predistorter. The estimation algorithm must be provided with signal data from the output of the RFPA $y_{RF}(t)$, which is acquired and digitized by the transmitter observation receiver (TOR) and associated analog-to-digital converters (ADC) to yield $y(n)$. The set of coefficients, once identified, are then used to update the predistortion engine that synthesizes the predistorted signal $x_{PD}(n)$ from the undistorted signal $x(n)$ using an appropriate mathematical formulation implemented with digital circuitry (adders, multipliers, etc.). The digital predistorted signal is then sampled by the digital-to-analog converters (DAC)

and then modulated onto a radio frequency carrier signal to generate the predistorted input to the RFPA $x_{PD,RF}(t)$. The qualifier 'digital' in digital predistortion refers specifically to the fact that the predistorted signal is synthesized in the digital domain. Predistortion can be realized using both direct and indirect learning (also known as the model reference adaptive control and self-tuning regulator approaches, respectively), which involve distinct architectures [7, 8]. The indirect learning scheme (Figure 7) uses the input and output of the RFPA to generate a reverse (post-inverse) model of the RFPA. The coefficients of this post-inverse model are copied over to the predistorter, verbatim. For the training of the post-inverse, the error being minimized is the difference between the output of the reverse model $z(n)$, and the input to the RFPA $x_{PD}(n)$; the operative assumption is that when the post-inverse is fed with the output of the RFPA, it will ideally recreate the input of the RFPA.

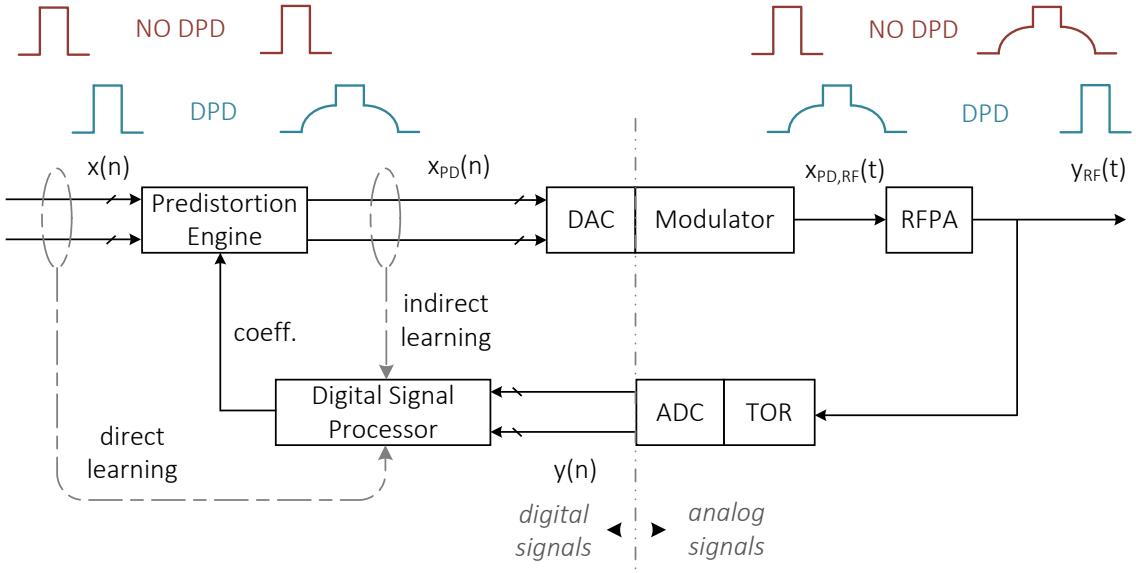

Figure 6: Block diagram of a digital predistorter for a single-band transmitter

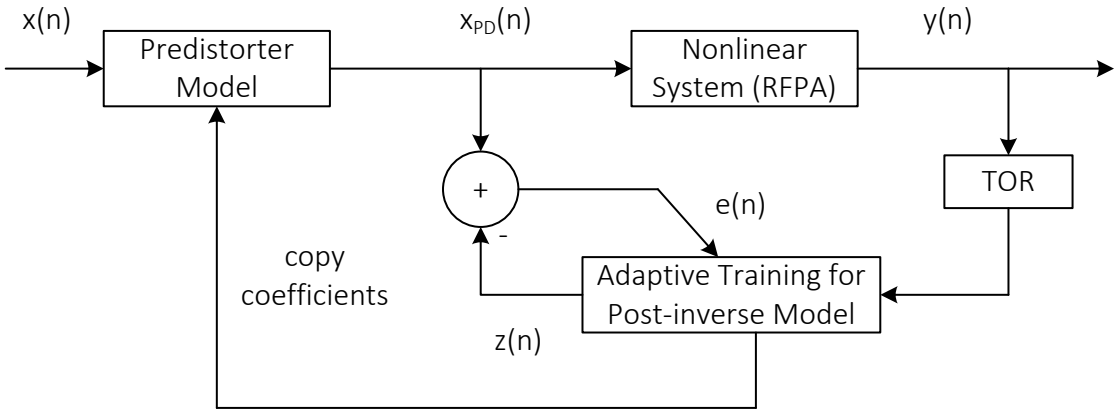

Figure 7: Indirect learning architecture for DPD

Opponents of the indirect learning approach argue that, because the RFPA and its inverse are non-commutative, identifying a post-inverse of the RFPA and placing a copy of it before the RFPA does not fully negate distortions. This has led to the proposal of direct learning as an alternative (Figure 8), in which the pre-inverse of the RFPA is identified and used as the predistorter. For the training of the pre-inverse, the error being minimized is the difference between the output of the RFPA $y(n)$, and the input $x(n)$ of the predistorter; the operative assumption here is that when the RFPA is cascaded after its true pre-inverse there will be no difference between the normalized output $y(n)$ and input $x(n)$ of the system.

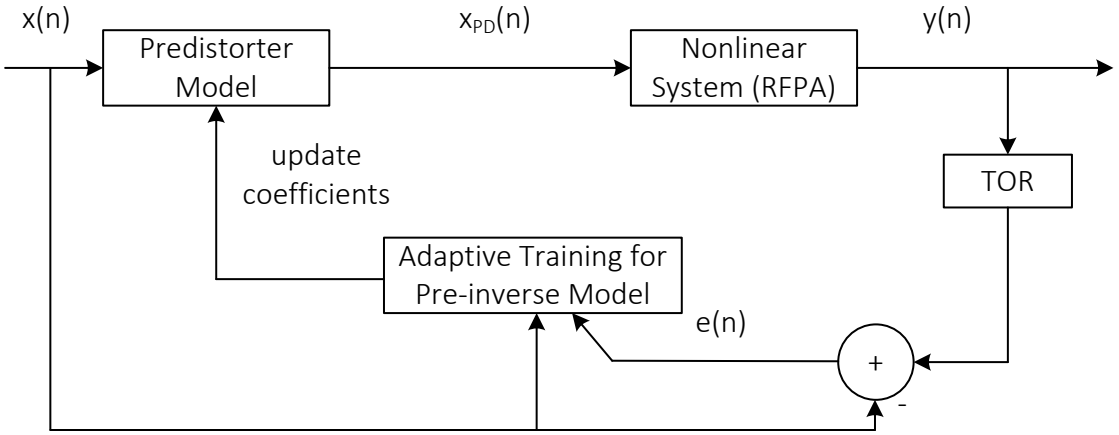

Figure 8: Direct learning architecture for DPD

The indirect learning method is conceptually easier to understand, because it attempts to identify a RFPA post-inverse model directly from the observed $y(n)$ and $x_{PD}(n)$ signals, i.e. some closed-form function that maps $y(n)$ to $x_{PD}(n)$,. In the case of direct learning, an equivalent function in terms of $x(n)$ and $y(n)$ cannot be formulated, as $y(n)$ is the result of a cascade of the predistorter and the RFPA, and the model for the latter is never identified. Essentially, indirect learning attempts a faithful construction of the RFPA post-inverse by reversing its transfer characteristic, while direct learning arrives at the RFPA pre-inverse while being agnostic to its transfer characteristic. Since this thesis is concerned with the proposal and validation of a new predistorter formulation, the conceptually simpler and more widely adopted approach of indirect learning has been used throughout. However, it is acknowledged later in the thesis that direct learning is necessary to truly realize a low-cost analog predistorter due to the architectural implications of the indirect approach.

Among various predistortion formulations, one of the most comprehensive ones is the Volterra series [25], which has been used extensively in the modeling of physiological systems, satellite communication links and microwave circuits. While is efficacy in modeling RFPAs is undisputed, the computational complexity of the Volterra series has discouraged its adoption into real-time, low-power applications where processing power is limited. In recognizing that all terms (or kernels) of the Volterra series are not equally critical in modeling, successful

attempts have been made to reduce its complexity, through the ad-hoc pruning of kernels [26], through a priori pruning based on the knowledge of physical mechanisms by which RFPAs non-idealities are generated [27], and even by directly deriving a complexity reduced formulation from a physically inspired multi-block model of the RFPA [12]. Further simplifications to the Volterra series have been devised, which significantly compromise modeling efficacy in return for simplicity. The memory polynomial [9] is the most prevalent of these, as it provides a moderate compromise between complexity and performance. The envelope memory polynomial is a simpler variant of the memory polynomial that discards linear memory terms and only models nonlinear memory effects [10]. Yet another class of formulations are the Hammerstein and Wiener [28], which decouple nonlinearity and memory terms into either a cascade of a memoryless nonlinearity followed by a linear / weakly nonlinear memory filter (Hammerstein / Augmented Hammerstein [29]) or the reverse (Wiener). Disposing the memory terms entirely results in the memoryless or static polynomial [30], which is decidedly on the simple end of the spectrum and hardly appropriate for contemporary linearization challenges. Equations 2.3 and 2.4 show the expressions for the Volterra series and static polynomial, respectively. In the equations below, M and m are the memory depth and index, respectively, N and k are the nonlinearity order and index, respectively, and h_k and a_k are the complex-valued coefficients. Figure 9 illustrates the relative positions of these formulations in terms of their "complexity" and their modelling capability, which are always at odds. Complexity in this case refers to mathematical complexity, and can be coarsely assessed by the number of complex multiplications needed to realize the predistortion function. At the top-end of modelling capability, the complexity C_{VOL} of the Volterra series is a function of its nonlinearity order N and memory depth M (assumed uniform for all kernels) and quickly becomes unmanageable as N and M grow, as shown by Eq. 2.5. In contrast, the complexity C_{STAT} of the memoryless polynomial is only a function of its nonlinearity, as in Eq. 2.6, and allows for easy implementation.

$$y_{VOL}(n) = \sum_{m_1=0}^{M-1} h_1(m_1)x(n-m_1) + \sum_{m_1=0}^{M-1}\sum_{m_2=m_1}^{M-1} h_2(m_1,m_2)x(n-m_1)x(n-m_2) + \dots \\ \sum_{m_1=0}^{M-1}\dots\sum_{m_N=m_{N-1}}^{M-1} h_N(m_1,\dots m_N)x(n-m_1)\dots x(n-m_N) \tag{2.3}$$

$$y_{STAT}(n) = \sum_{k=1}^{N} a_k x^k(n) \tag{2.4}$$

$$C_{VOL}(N,M) = \sum_{n=1}^{N} \frac{(M-1+n)!}{(n-1)!(M-1)!} \tag{2.5}$$

$$C_{STAT}(N) = N \tag{2.6}$$

$$M, N \in \mathbb{Z},\ a_k, h_k \in \mathbb{C}$$

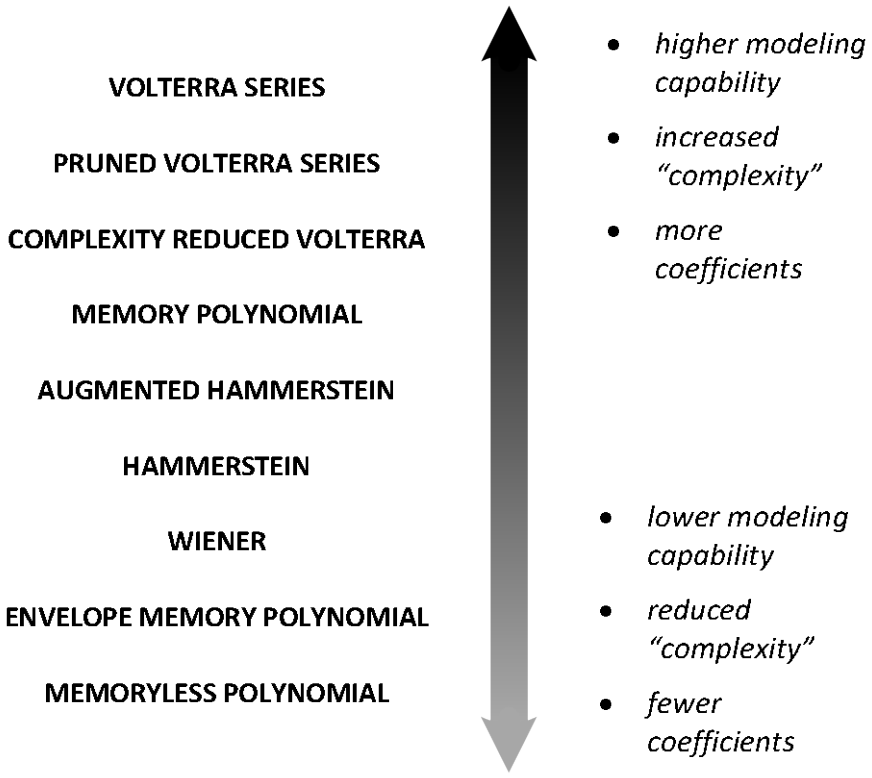

Figure 9: Relative complexity and capability of predistorter formulations

While the choice of formulation determines the theoretical accuracy with which an RFPA can be linearized, a far more practical concern is the accurate identification of the optimal coefficients that bring the model performance close to its theoretical limit. It is the accuracy of coefficient estimation, rather than model capability, that usually becomes the limiting factor in linearization success, as will be noted in later chapters. All the formulations previously referenced, from the Volterra series to the memoryless polynomial, share a common property that greatly lessens the burden of coefficient identification – they are linear functions of the unknown coefficients. This property allows the use of the least squares estimation (LSE) technique (i.e., linear regression) to estimate the optimal coefficients. In the indirect learning architecture, for example, application of least squares requires two sets of $1 \times L$ data vectors, X and Y, where Y is comprised of data points from the digitized output signal of the PA and X is comprised of corresponding data points from the input signal to the PA, and L is the length of the vectors, usually chosen to be 10,000 to ensure a representative sample. For reverse modelling, a matrix A, comprised of m basis vectors determined by the choice of formulation, is generated from Y which, when multiplied with the vector of unknown coefficients P, should ideally reconstruct X (for reverse modelling). The unknown P is then

determined using the Moore-Penrose pseudoinverse of A, $A^{\dagger}$ as shown in Eq. 2.7 (matrix dimensions are shown in subscripts).

$$\begin{aligned} &A_{L\times m} = f(Y_{L\times 1}) \\ &A_{L\times m}P_{m\times 1} = X_{L\times 1} \\ &P_{m\times 1} = A_{m\times L}{}^{\dagger}X_{L\times 1} = (A_{m\times L}{}^{T}A_{L\times m})^{-1}A_{m\times L}{}^{T}X_{L\times 1} \end{aligned} \tag{2.7}$$

Using linear regression guarantees that the identified coefficients are truly optimal in a least square sense, and allows the data-fitting capability of a formulation to be assessed. However, the LSE approach involves the storage of massive data vectors, inversion of an $m \times m$ matrix, as well as $m \times L$ complex multiplications – requiring computational resources and memory that are prohibitive to low-power, real-time applications. The less computationally intensive alternative to LSE is the Recursive Least Squares (RLS) technique which recursively estimates the coefficients. The RLS solver must be supplied with an initial guess of the coefficients. At each step, the solver updates its guess with a correction that is equal to the product of a calculated gain, and the error in predicting the current observation with the current guess.

The prime advantage of using RLS is that it does not need to store or manipulate large vectors of inputs and observations – after each update to the coefficients, the input and observation points used in the calculation of the update are discarded. Furthermore, an RLS algorithm can track changes in the operating conditions of the PA on a sample-by-sample basis and update the coefficients of the predistorter in real-time whereas the LSE technique recalculates the coefficients every time it is applied, and it can only be applied once enough samples of the input and output of the PA have been aggregated.

The caveat with RLS is that, unlike the LSE which yields the truly optimal coefficients, it converges to the vicinity of the optimal coefficients, with subsequent updates causing the solution to fluctuate around the optimal. It is possible to make the solver converge closer to the optimum by scaling down the step-size of the correction, but the added accuracy comes at the cost of a slower convergence rate, resulting in higher cumulative complexity due to repeated matrix inversions.

LSE and RLS are archetypal instances of block-wise vs. recursive estimation methods, and attempts to reduce their complexity and/or improve their accuracy have been the subject of several publications, a comprehensive assessment of which is beyond the scope of this thesis. Nonetheless, recursive and real-time estimation methods remain a promising approach for practical implementations of predistortion solutions.

Ideally, an accurate identification of the reverse PA model would result in a perfect reconstruction of the PA input from the PA output. Therefore, a common approach to assess the optimality of an identified predistorter model is to feed it with all of the points comprising the data-vector Y (the normalized output of the PA used for identification), and measured how accurately it re-constructs the data-vector X (the normalized input of the PA used for

identification). The 'goodness' of fit between the actual and reconstructed data (x_i vs $\hat{x}_i$) is measured using the normalized mean squared error (NMSE, Eq. 2.8). NMSE values below – 40 dB usually indicate an acceptable degree of accuracy for predistortion applications.

$$NMSE = 10\log_{10}\left(\frac{\sum_{i=1}^{L}\left|\hat{x}_i - x_i\right|^2}{\sum_{i=1}^{L}\left|x_i\right|^2}\right) \tag{2.8}$$

2.3 Alternatives to Digital Predistortion

Several emerging trends in the field of wireless communications are beginning to challenge the utility of digital predistortion as means of RFPA linearization. Future wireless networks are envisioned to feature greater number of micro, pico, and femto-cells, which will be situated strategically to augment localized weak-spots within the large coverage area of the base-stations in order to provide seamless and uniform network connectivity regardless of location. Because of their much smaller coverage area, these small cells will need physically smaller amplifiers capable of transmitting at watt or sub-watt levels at the most. Interestingly, the degree of non-idealities that need to be corrected for in small cells will change very little compared to macro-cells, for the following reason: nonlinearity and memory effects are by-products of design choices made to maximize RFPA efficiency (such as the choice of biasing, topology, and matching networks) and there is no reason to believe that efficiency will be any less of a concern for small cells, especially with growing pressure on the ICT industry to reduce its carbon footprint. Consequently, the corrective capabilities of linearization techniques such as DPD are not expected to change through the transition from macro-cell to small-cell. For base-station macro-cells, the power overhead incurred by DPD (in the order of watts) was a relatively small price to pay for being able to transmit hundreds of watts using efficiently designed RFPAs. In the case of small-cells, the DPD power consumption will not scale down (since the requirement for linearity is unchanged), and will be a massive price to pay for being able to transmit a couple of watts efficiently. This motivates seeking alternative predistortion techniques that perform comparably to DPD without consuming as much power.

Proposing a low-power alternative necessitates understanding the factors responsible for the high power consumption in DPD. The reader is referred back to Figure 6 of the previous section, which showed the architecture of a typical single-band transmitter with a DPD system to linearize the RFPA.

The primary source of power consumption is the baseband digital circuitry, particularly the DACs preceding the modulator. In order to maintain the fidelity of the signal through the conversion process, the sampling rate of the DAC must at least be twice as high as the highest frequency component of the baseband signal, as per the Shannon-Nyquist theorem. Unfortunately, higher sampling rates equate to higher power consumption in DACs, and the problem is aggravated by digital predistortion, which causes the baseband signal to experience

a $\times 5$ bandwidth expansion [31]. Thus, an undistorted WCDMA signal of 20 MHz bandwidth will roughly occupy 100 MHz of bandwidth after predistortion, and will need to be sampled by two 200 Ms/s DACs (for the in-phase and quadrature phase components). Commercial DACs operating at this sampling rate, even low-resolution ones [32], will consume hundreds of milliwatts alone – and this is over and above the power consumption of the digital predistortion circuitry, which will need to be clocked at the same rate. Another major source of power consumption is the transmitter observation receiver which employs ADCs to digitize the distorted output signal of the RFPA and provide it to the DPD engine to train the predistorter. Much in the same fashion as the predistorter, the RFPA causes the transmitted signal to experience a bandwidth expansion that is detrimental to the power consumption of the ADCs. For small cells transmitting at watt and sub-watt levels, these sources of power dissipation easily negate the improvement in efficiency offered by linearization. The problem will only worsen as communication signals experience a widening of modulation bandwidth to support increasingly high data-rate applications. LTE-A is expected to provide support for contiguous carrier-aggregated signals with modulation bandwidths reaching up to 100 MHz [33], while 5G air interfaces are expected to support modulation bandwidths in the range of several hundred MHz to a GHz [34], underscoring the need for alternative predistortion architectures where power consumption can be managed more reasonably with respect to widening modulation bandwidth. Analog predistortion is a strong contender in this regard, and the next chapter develops a formulation that is suited for an APD architecture.

Chapter 3
Single-band Analog Predistortion

This chapter is organized as follows. Section 3.1 discusses architectural differences between DPD and APD, and how they translate to savings in power consumption for APD while creating additional challenges in the design of the predistorter engine. Section 3.2 discusses compatibility issues between the APD architecture and the majority of pre-existing DPD formulations – with the exception of the envelope memory polynomial. Limitations of the EMP for wideband linearization are identified, and subsequently addressed by proposing a cascaded FIR-EMP formulation. In Section 3.3, the challenge of identifying the proposed formulation's coefficients is addressed using two possible identification schemes, and in Section 3.4, its performance is validated using an experimental setup involving a physical RFPA excited with single-band, wide bandwidth communication signals.

3.1 Transition from Digital to Analog

Of the prior works that served as prototypical demonstrators of single-band APD, a brief discussion of which has already been presented in section 2.2, [17] achieves predistortion of narrowband signals with unprecedented low power by directly predistorting the RF signal using an analog engine as opposed to conventional DPD approaches which predistort the baseband signal using a digital engine. The transitions from digital to analog, and baseband to RF, strongly influence the constituents and architectural layout of the predistorter. The architecture in turn imposes limitations on the type of predistorter formulation that can be synthesized. In this chapter, these limitations are examined in order to construct a new predistorter formulation that is compatible with the APD architecture and boasts robust performance in wideband contexts. Challenges of identifying the predistorter's coefficients are discussed as well. The proposed formulation (and method of identification) is evaluated on its ability to linearize a physical RFPA transmitting single-band wide bandwidth signals and compared against other common predistortion formulations.

Figure 10 depicts the architecture of a single-band analog predistorter that has been integrated into a transmit path containing the RFPA. Contrasting this with the DPD architecture of Figure 6 reveals how the predistortion engine has relocated from the digital to the analog domain, where low power circuitry can employed to generate the predistortion correction signals $I_c(t)$ and $Q_c(t)$, which are then administered directly to the undistorted signal $x_{RF}(t)$ to generate the predistorted signal $x_{PD,RF}(t)$ which is fed into the PA. As a consequence of this relocation, the baseband signal $x(n)$ is still undistorted prior to up-conversion. This plays a major role in reducing power consumption as the baseband DACs, which no longer need to accommodate the $\times 5$ bandwidth expansion caused by predistortion and can be clocked at much lower speeds.

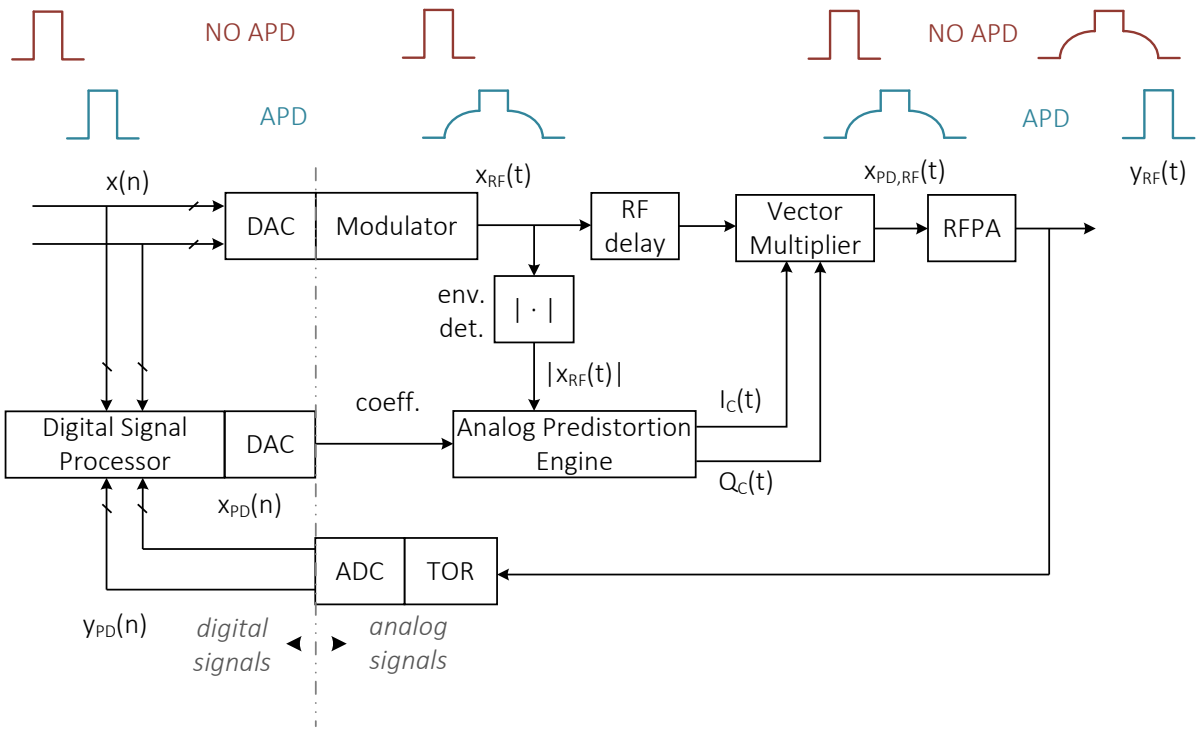

Figure 10: Block diagram of an analog predistorter for a single-band transmitter

The caveat of this relocation is that the predistorter engine no longer has access to the digital in-phase and quadrature-phase signals that carry both the amplitude and phase information of the undistorted signal, which restricts the predistorter output to be a function of just the undistorted signal amplitude or envelope (sensed by the envelope detector). As a benefit, because the signal envelope varies with the frequency of baseband and not the frequency of the carrier, the analog predistorter engine does not need to operate at radio frequencies. However, the amplitude-only restriction severely limits the efficacy of the predistorter for wideband signal transmission. Trivially, this limitation could be avoided by making both the baseband in-phase and quadrature phase signals available to the predistorter using two approaches, but neither of them are feasible. One approach would be use a demodulator (quadrature coupler, mixer, local oscillator) in place of an envelope detector to re-obtain the baseband signals from the RF, which is severely wasteful in terms of cost, size and power consumption. The other approach is to directly give the predistorter engine access to the in-phase and quadrature-phase baseband signals; this precludes having the elegance of an RF-in-RF-out predistorter and creates two divergent paths for the baseband signal, (one through the modulator and another through the engine) both of which must be carefully co-designed to reduce path delay and gain mismatch – such design complications are best avoided. Another caveat of the relocation is that the digital signal processor that trains the predistorter coefficients no longer has access to the predistorted input of the RFPA, since the training occurs in digital baseband and the predistortion in the analog pass band (compared to DPD, where both training and predistortion occur in the same domain). A second TOR would be required to convey RFPA input data to the training engine, which would defeat the objective

of reducing power overhead. Realistically, the digital signal processor (DSP) in Figure 10 would need to implement direct learning in the absence access to the RFPA input. While this thesis validates the proposed predistorter formulation with indirect learning as a first step, the importance of eventually transitioning to direct learning has been stressed in the discussion of future work.

A distinct feature of the APD architecture also differs significantly in that it uses a vector multiplier to adjust the amplitude and phase of the undistorted RF signal to obtain the predistorted RF output, based on the correction signals $I_c(t)$ and $Q_c(t)$. These correction signals are synthesized by an analog engine that operates on the envelope of the undistorted signal, which carries only amplitude information. Furthermore, to account for the finite time needed by the analog engine to synthesize, a deliberate delay element must be introduced into the path of the undistorted signal to ensure that it is synchronized with the correction signals. In contrast, the DPD engine directly operates on the in-phase and quadrature-phase components of the undistorted baseband signal (which carry both amplitude and phase information) and generates the predistorted baseband output directly without needing delay synchronization.

APD will also introduce challenges specific to analog integrated circuit design. A digital predistorter can be implemented on commercial field programmable gate arrays (FPGA's) by programming the appropriate multiplication and addition operations to be carried out using a hardware description language. For APD, specific predistorter formulations must be implemented with custom-designed integrated circuits to perform operations of multiplication and addition on analog voltage/current signals. Designers must worry about (i) the impact of circuit organization and layout on thermal noise and delay mismatch, (ii) the limited dynamic range of the circuits which can cause signals to be clipped or driven into the noise floor, and (iii) the effect on accuracy of unavoidable circuit non-idealities such as signal offset and device mismatch. These concerns grow exponentially as the predistorter formulation becomes more complicated and the number of mathematical operations, complex multiplications in particular, increases. Consequently, not all predistorter formulations are well suited to the transition from digital to analog. The Volterra series, for example, is poorly suited to APD despite its tremendous linearization capacity. The design of an analog Volterra predistorter would be intractable due to the complexity of the formulation, besides which, any theoretical improvements in modelling accuracy that it could offer compared to simpler formulations would be offset by the cumulative impact of analog circuit non-idealities. Selecting a formulation that only retains essential mathematical terms while being compatible with the APD architecture is necessary, as will be discussed in the coming section.

It should also be noted that while the APD architecture addresses the issue of DAC power consumption due to the bandwidth expansion caused by predistortion, it has no impact on the ADC power consumption in the TOR. The issue will not be addressed in this thesis, since independent research efforts are underway to limit this power consumption by reducing the

'observation bandwidth' of the TOR and thus limiting the sampling rate and clocking speed of its associated circuitry [35].

3.2 Analog-friendly Predistorter Model

Of the numerous predistorter formulations reviewed in Chapter 2, very few are compatible with analog predistortion, because of the restrictions imposed by its architecture and hardware design challenges. A central limitation arises from its use of the vector multiplier, which applies the predistortion as a correction to the amplitude and phase of the undistorted signal. The vector multiplier operation is depicted in Figure 11.

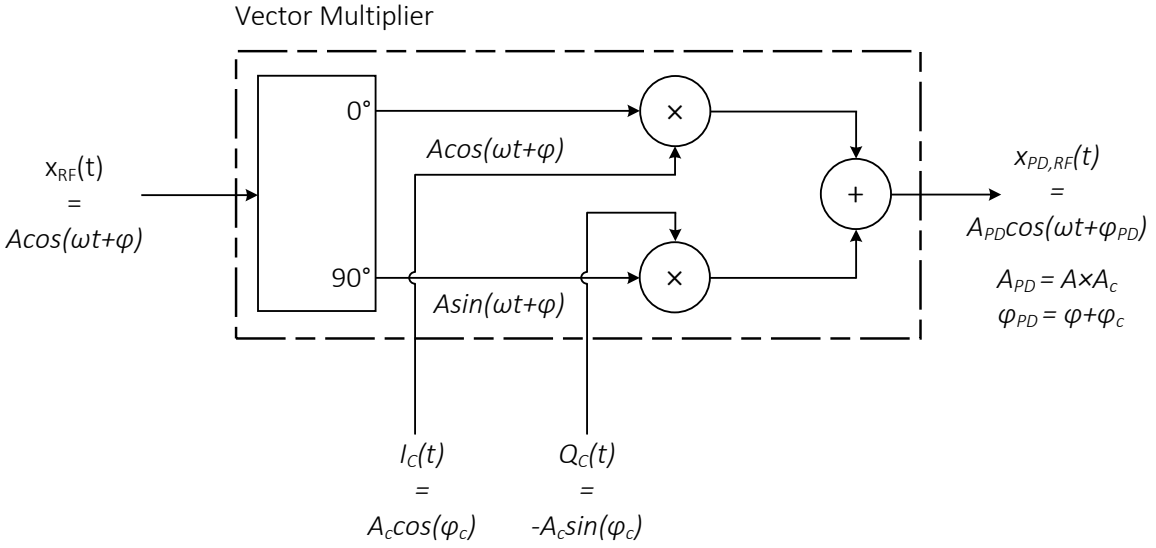

Figure 11: Operation of the vector multiplier

$$
\begin{aligned}
&A_{PD}(t)\cos\left(\omega t+\varphi_{PD}(t)\right)=A(t)\times A_c(t)\cos\left(\omega t+\varphi(t)+\varphi_c(t)\right)\\
&\Rightarrow \mathrm{Re}\left\{A_{PD}(t)e^{j\varphi_{PD}(t)}e^{j\omega t}\right\}=\mathrm{Re}\left\{A(t)e^{j\varphi(t)}\times A_c e^{j\varphi_c(t)}e^{j\omega t}\right\}\\
&\Rightarrow \mathrm{Re}\left\{\left(x_{PD,I}(t)+jx_{PD,Q}(t)\right)e^{j\omega t}\right\}=\mathrm{Re}\left\{\left(x_I(t)+jx_Q(t)\right)\times\left(I_c(t)+jQ_c(t)\right)e^{j\omega t}\right\} \qquad (3.1)\\
&\text{where } x_{n,I}=A_n\cos\varphi_n \text{ and } x_{n,Q}=A_n\sin\varphi_n
\end{aligned}
$$

Eq. 3.1 shows how the vector multiplication with RF signals is equivalent to a baseband operation in which the predistorted output $x_{PD}(t)$ is obtained by multiplying the undistorted input $x(t)$ with a corrective 'gain' that is equal to $I_c(t)+jQ_c(t)$. Thus, for a predistorter formulation to be compatible with the APD architecture, it has to be possible to express its output – the predistorted signal – as a product of its input – the undistorted signal – and a corrective gain. Furthermore, the corrective gain must only be a function of the envelope of the undistorted signal, as the analog predistortion engine which synthesizes the gain only senses the amplitude through the envelope detector and not the phase. The majority of predistortion formulations do not satisfy this criterion, but it is easy to show that the envelope memory polynomial does. Eq. 3.2 shows how the time-domain analog equivalent of the

discrete representation of the EMP can be re-arranged into the needed form. A similar rearrangement can be applied to the static polynomial to show that it is compatible with APD as well. However, the well-known memory polynomial cannot be expressed in the desired form, as shown in Eq. 3.3, because it includes certain memory terms of the form $x(t-m\tau)\left|x(t-m\tau)\right|^k$ that cannot be expressed as a function of $x(t)$. The Volterra series, its inherent complexity notwithstanding, is also unsuitable for APD for the same reason. Given the limited choice between either the static polynomial or envelope memory polynomial, and the modelling advantages of the latter, the candidate for an APD formulation becomes obvious.

$$\begin{aligned} x_{PD,EMP}(t) &= \sum_{m=0}^{M}\sum_{k=0}^{N-1} a_{m,k}\, x(t)\left|x(t-m\tau)\right|^k \\ &= x(t)\times\left\{\sum_{k=0}^{N-1} a_{0,k}\left|x(t)\right|^k\right\} + x(t)\left\{\sum_{k=0}^{N-1} a_{1,k}\left|x(t-\tau)\right|^k\right\} + \cdots x(t)\left\{\sum_{k=0}^{N-1} a_{M,k}\left|x(t-M\tau)\right|^k\right\} \\ &= x(t)\times\left\{\sum_{m=0}^{M}\sum_{k=0}^{N-1} a_{m,k(\mathrm{Re})}\left|x(t-m\tau)\right|^k + j\sum_{m=0}^{M}\sum_{k=0}^{N-1} a_{m,k(\mathrm{Im})}\left|x(t-m\tau)\right|^k\right\} \\ &= x(t)\times\left(I_c(t)+jQ_c(t)\right) \end{aligned} \tag{3.2}$$

$$\begin{aligned} x_{PD,MP}(t) &= \sum_{m=0}^{M}\sum_{k=0}^{N-1} a_{m,k}\, x(t-m\tau)\left|x(t-m\tau)\right|^k \\ &= x(t)\left\{\sum_{k=0}^{N-1} a_{0,k}\left|x(t)\right|^k\right\} + x(t-1)\left\{\sum_{k=0}^{N-1} a_{1,k}\left|x(t-1)\right|^k\right\} + \cdots x(t-M\tau)\left\{\sum_{k=0}^{N-1} a_{M,k}\left|x(t-M\tau)\right|^k\right\} \end{aligned} \tag{3.3}$$

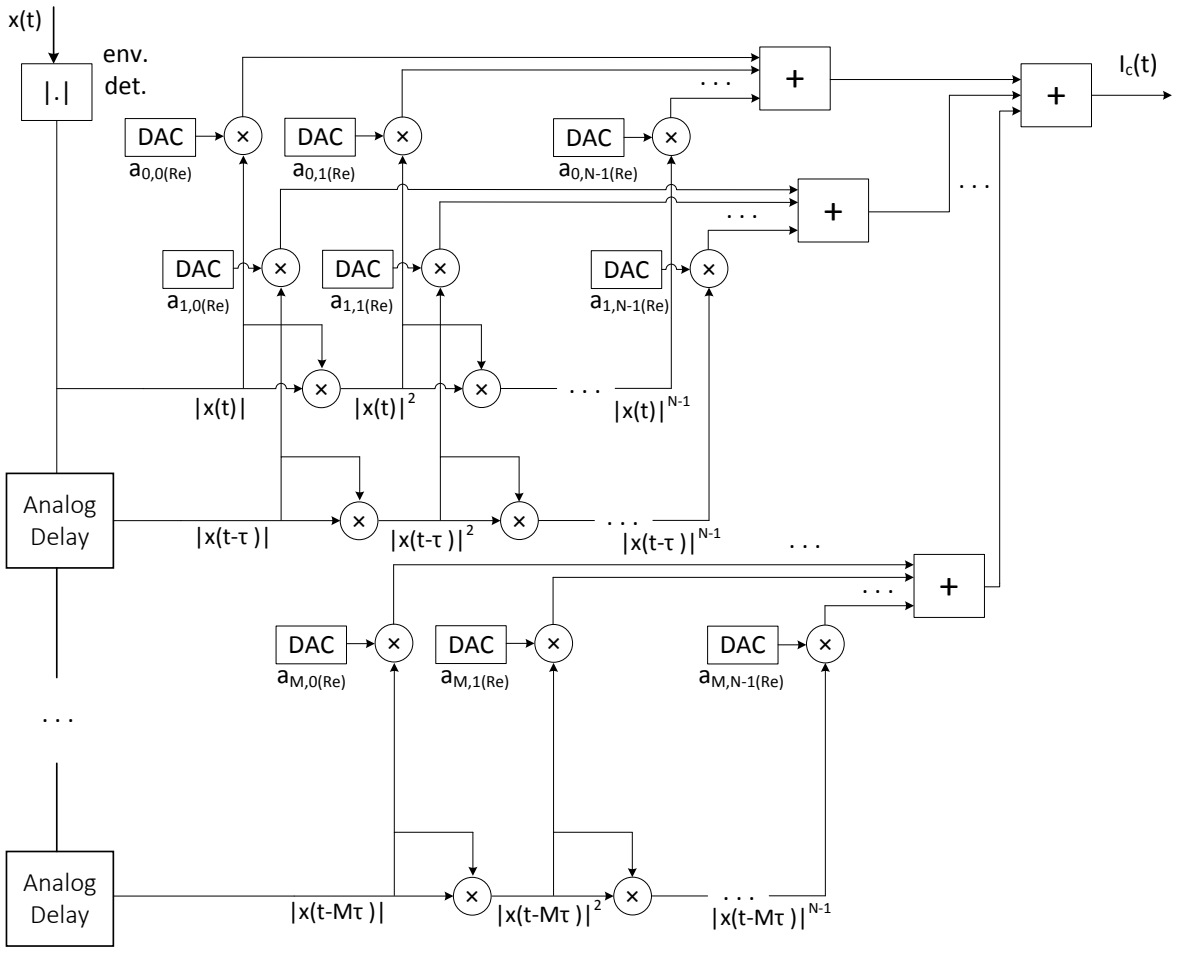

Figure 12: Block diagram of a single-band analog predistorter engine

The schematic of an analog predistortion engine to realize the EMP formulation is presented in Figure 12. The figure shows the multipliers, adders, and DACs necessary to realize the correction signal $I_c(t)$. An identical structure would be used to generate $Q_c(t)$, except with the DACs being used to generate the imaginary part $a_{m,k(\mathrm{Im})}(t)$, of the complex coefficients. The correction signals do not possess terms of the form $x(t-m\tau)\left|x(t-m\tau)\right|^k$. These terms can only be synthesized by either (i) down-converting the RF signal to re-obtain the baseband signal, or (ii) enabling access to the baseband signal prior to its up-conversion, neither of which are feasible for reasons already discussed. In [10], where EMP was proposed, discarding these problematic memory terms does not appear to have any significant impact on modelling performance, and the EMP and MP appear comparable at first glance. Further investigation reveals that the performance of the two formulations diverges as the modulation bandwidth of the test signal is increased – the EMP struggles to match the performance of the MP. The difference is evident from Figure 13, which compares the spectra of progressively wider bandwidth signals linearized by MP (black) against EMP (blue) and no predistortion (red). While the linearized spectra are identical in the case of very narrow modulation

bandwidths, at 20 MHz the ACLR of the EMP-linearized spectrum is about 10 dB worse than the MP-linearized spectrum, and the discrepancy continues to increase for wider bandwidths.

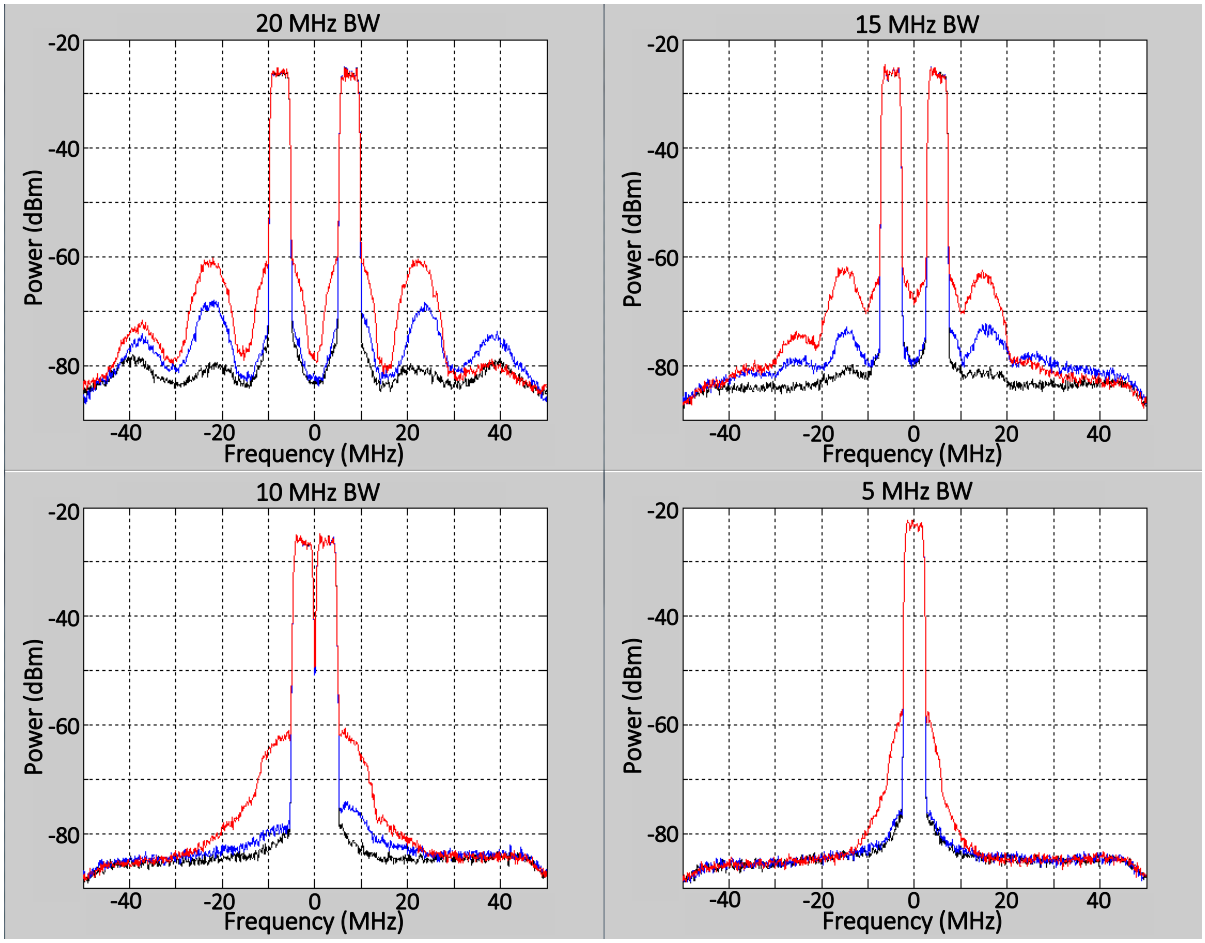

Figure 13: RFPA spectra linearized using MP (black), EMP (blue), and unlinearized (red)

An AM/AM and AM/PM plot generated using the predistorted signal can reveal how effective the predistorter formulations are in modelling the reverse gain and phase characteristics of the PA. Figure 14(a). and 14(b). show the AM/AM and AM/PM curves modelled by the MP and EMP formulations respectively, contrasted against the reference curve of the PA. While both MP and EMP are able to reasonably model the nonlinearity and dispersion of the RFPA transfer characteristic at high power levels (Figures 14(c). and 14(d).) where nonlinear memory effects dominate, the EMP is unable to model the linear memory effects of the PA that manifest at low signal power, as evidenced by the lack of dispersion in the teal-colored curves. It is well understood that modulated signals that occupy a larger frequency bandwidth are more sensitive to frequency dispersion and will stimulate more memory effects in RFPAs; the inability of the EMP to model linear memory effects would explain why it performs progressively worse at higher bandwidths.

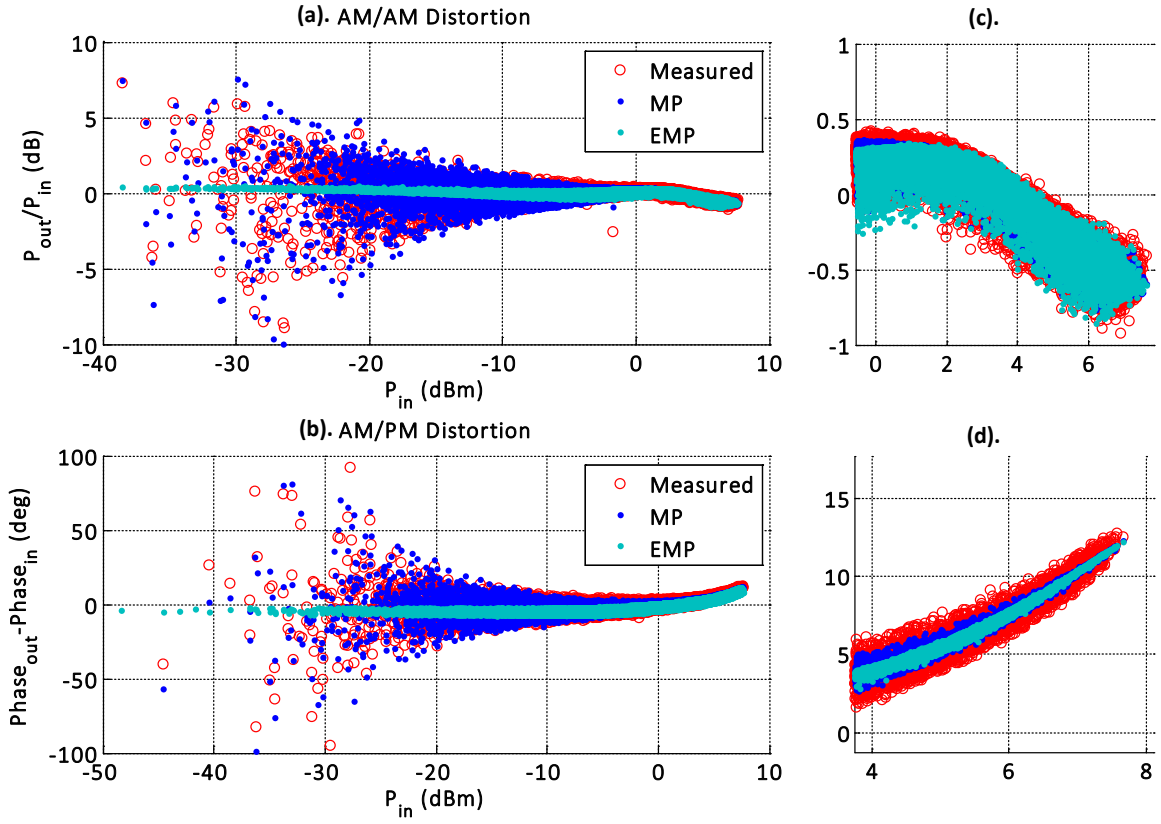

Figure 14: Contrast between AM/AM and AM/PM modeled using the MP and EMP

Unfortunately, the limitation of the EMP is attributable to the same property that makes it a viable candidate for APD, namely the absence of terms $x(t-m\tau)\left|x(t-m\tau)\right|^k$. Of particular significance are the terms associated with index $k=0$, i.e., linear memory terms of the form $x(t-m\tau)$ that can be generated by passing $x(t)$ through a finite impulse response (FIR) filter. These terms are essential to the modelling of linear memory effects that arise due to frequency dispersion caused the matching networks before and after the transistor. (FIR filters have been widely used to correct chromatic dispersion in optical communication systems as well [36, 37]); this was verified in MATLAB simulation by observing that the EMP formulation, when supplemented by the addition of terms of a finite impulse response filter to form a parallel structure (i.e., 'FIR||EMP' in Figure 15), performs as well as the full memory polynomial in modelling the reverse characteristic of an RFPA driven with wideband signals ranging up to 80 MHz. Nonetheless, it is not feasible to realize these FIR terms in an APD implementation because they are not functions of the signal envelope.

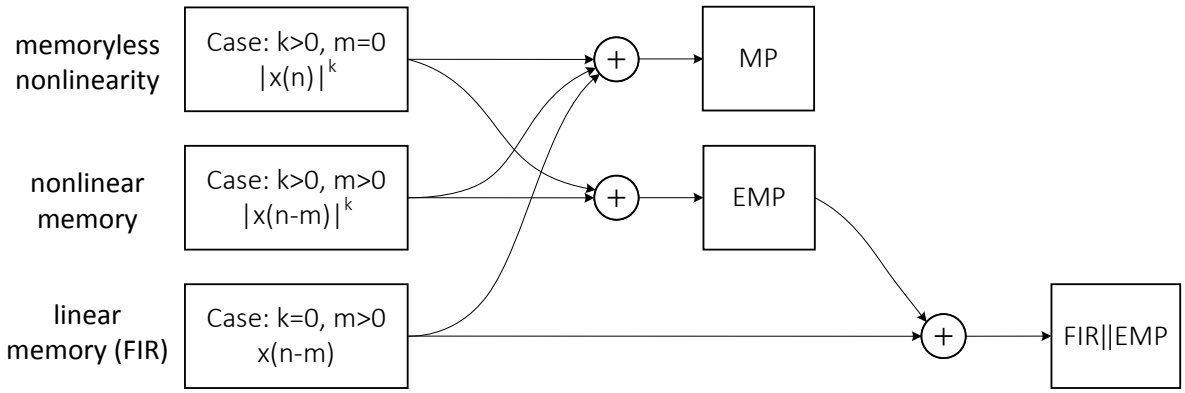

Figure 15: Relationships between the mathematical bases of MP, EMP, and FIR||EMP

As a solution to the above conundrum, an alternative arrangement is proposed, which cascades the FIR filter before the EMP formulation, instead of placing it in parallel. This re-arrangement has a specific advantage: it allows the FIR filtering to be realized effortlessly in the digital baseband domain using delay blocks, while allowing the EMP to be implemented using analog circuitry as previously proposed. Figure 16 shows the block diagram of the proposed FIR-APD scheme, which differs from the APD scheme of Figure 10 only by the inclusion of an FIR-filter implemented in digital. The resulting FIR-APD formulation is shown in Eq. 3.4, in which M and N are the memory depth and nonlinearity orders of the EMP, and V is the order of the FIR filter. Results in the measurement section confirm that, even in a cascade re-arrangement, the FIR filter is able to provide the much needed capability of modeling linear memory effects which, when combined the capability of the APD to model static nonlinearity and nonlinear memory effects, provides the essential bases for describing the non-ideal transfer characteristics of contemporary RFPAs. The proposed formulation bears some similarity to earlier cascaded models such as the Hammerstein and Wiener. It should be noted that the inclusion of the filtering action does not 'predistort' the baseband signal significantly to impact power consumption – thus the proposed FIR-APD architecture still retains the advantages of the APD architecture while promising an improvement in linearization performance.

$$x_{FIR-EMP}(t) = \sum_{m=0}^{M}\sum_{k=0}^{N-1} b_{m,k}\, x_{FIR}(t)\left|x(t-m\tau)\right|^{k}$$
$$x_{FIR}(n) = \sum_{v=0}^{V} a_v x(n-v) \qquad (t = nT) \tag{3.4}$$

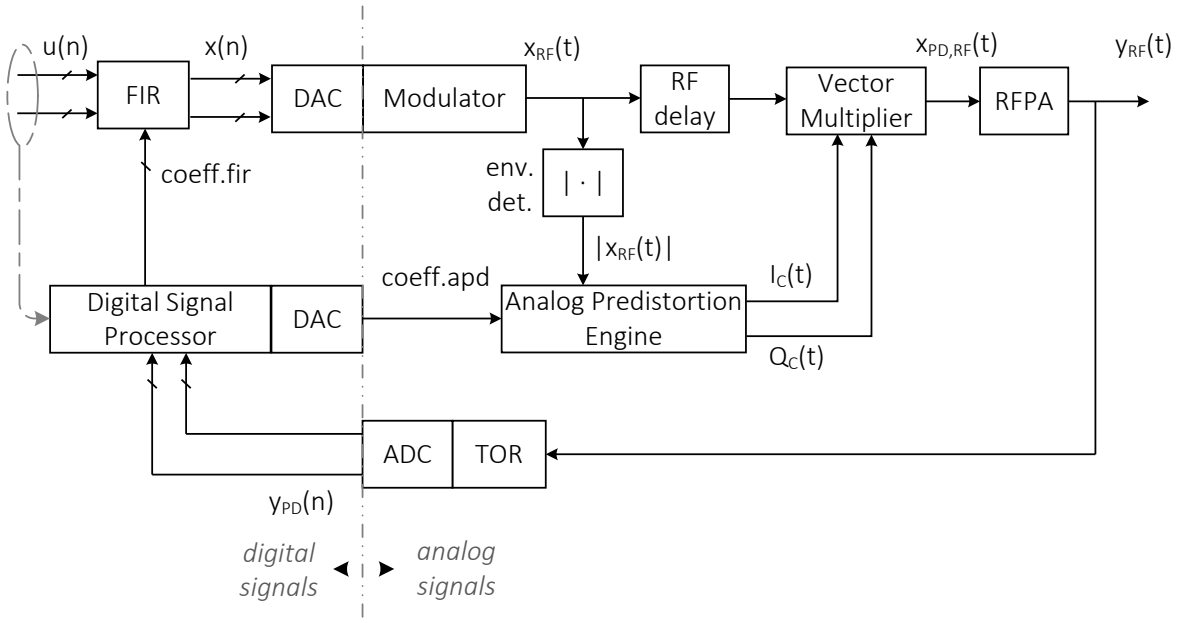

Figure 16: Block diagram of the FIR-APD system for a single-band predistorter

3.3 Analog Predistorter Model Identification

The discussion of predistorter model identification in Section 1 referred to the use of block-wise and recursive estimation techniques such as the pseudo-inverse based LSE and the RLS for determining the coefficients. Unfortunately, these commonly employed linear identification techniques are precluded from use with the proposed FIR-APD scheme because it does not satisfy the necessary condition of being linear with respect to the unknown coefficients. As illustrated with Figure 17 and Eq. 3.5, this issue is not unique to FIR-APD, and would arise for any predistorter formulation that is the composition two functions, in which the latter is nonlinear (such as the Wiener model).

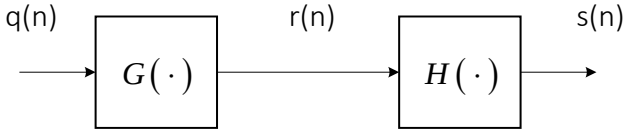

Figure 17: Cascading two functions to yield a composite function

$$G(q) = g_1 q(n) + g_2 q(n-1) \qquad H(r) = h_1 r(n) + h_2 r^2(n)$$

$$s = (H \circ G)(q) = \gamma_1 q(n) + \gamma_2 q(n-1) + \gamma_3 q^2(n) + \gamma_4 q(n) q(n-1) + \gamma_5 q^2(n-1) \tag{3.5}$$

$$\gamma_1 = h_1 g_1 \quad \gamma_2 = h_1 g_2 \quad \gamma_3 = h_2 g_1^2 \quad \gamma_4 = 2 h_2 g_1 g_2 \quad \gamma_5 = h_2 g_2^2$$

The issue can be addressed with recourse to nonlinear optimization; specifically, MATLAB's unconstrained nonlinear optimization function *fminunc*, was leveraged to demonstrate the highest modeling potential achievable with the proposed FIR-EMP formulation. *Fminunc* was

programmed to find the optimal coefficient vector c_{opt} that minimizes an objective function $f_{obj}(c)$ defined in Eq. 3.6. In the cost function of Eq. 3.6, $c = [a_v \;\; b_{mk}]^T$ is the concatenation of the coefficients of the EMP and FIR filter from Eq. 3.4, x_i are the training samples of the RFPA input, and $f_{FIR-EMP}(c, y_i)$ is the estimate of x_i yielded by Eq. 3.4 when it is evaluated with training samples of the RFPA output y_i.

$$f_{obj}(c) = \left(\frac{\sum_{i=0}^{L} \left| x_i - f_{FIR-EMP}(c, y_i) \right|^2}{\sum_{i=0}^{L} \left| x_i \right|^2} \right) \tag{3.6}$$

The objective function above essentially calculates the normalized mean square error that was presented in Eq. 2.8, sans logarithmic conversion. Minimization of this function will ensure that the resulting coefficients provide a good fit between the measured and modelled data, provided the training samples are chosen to be a representative subset of the entire signal. Usually, it is sufficient to choose training samples consisting of 10% of the entire set of data-points comprising x and y. The samples are also chosen from around the peak regions of the input and output signals, where the most pronounced nonlinear behavior is present, to ensure that the predistorter model captures the full extent of the nonlinearity.

It should be noted that $f_{FIR-EMP}(c, y_i)$ is a nonlinear function of the coefficients, owing to the composite nature of the proposed formulation, which motivates the use of *fminunc* in the first place. *Fminunc* itself can employ several different algorithms for optimization, which include the Quasi-Newton [38], Nelder-Mead [39], and Trust-Region [40] algorithms. The trust-region algorithm requires a user-specified gradient function for the predistorter formulation, which in the case of FIR-APD, entails determining partial derivatives of a tedious composite function with respect to each of its unknown coefficients. The simplex method of Nelder-Mead algorithm, which neither requires nor employs any gradient information, is generally less efficient. The Quasi-Newton method, which requires no user-specified gradient function but approximates it using the observed behavior of the objective function, was chosen for this work. Even though the coefficients yielded by this process cannot guarantee a global minimum in the NMSE, measurement results in the next section will demonstrate that with these coefficients, the envelope memory polynomial can outperform the memory polynomial from which it was derived.

As is true for any iterative optimization problem, the choice of an initial guess can affect the convergence to an optimal solution. To avoid prohibitively long search times, and to reduce the chances of converging on sub-optimal minima, a reasonable first guess for the coefficients can be constructed as follows. A parallel FIR||EMP model can be assumed for the predistorter, which is linear in the coefficients and has been shown to model the reverse transfer characteristic well. The coefficients of this model are identified using least squares estimation,

as described in Section 1. The EMP coefficients of this model are used as part of the first guess, while its parallel FIR coefficients are discarded and a cascade FIR filter with unity gain is used to complete the first guess (i.e. the first coefficient of the FIR filter is assumed 1, and the rest, zero). This is just one of many possible approaches at arriving at an initial guess; another approach could involve identifying a Hammerstein model as a first step, using its linear memory coefficients as the first guess for the FIR filter, and using its static polynomial coefficients as the first guess for the static part of the EMP (while setting its other coefficients to zero).

While the recourse to nonlinear algorithms allows the FIR-APD formulation to be evaluated to its fullest potential against other DPD formulations, it is not a practically viable solution, as it demands tremendous computational resources that cannot be provided by real-time, on-chip digital signal processing solutions for small cell applications. The *fminunc* implementation in MATLAB attempts to minimize the objective function by iteratively varying the choice of the coefficients using a line-search (where the search direction is informed by the Quasi-Newton approach chosen above). Each iteration involves a costly evaluation of the objective function, and the computational burden can grow quickly as the number of coefficients increase with higher model order – leading to excessively long computation times or prohibitive requirements in processing power. For an FIR-EMP model with a modest 25 complex coefficients, for example, the above approach would involve a search vector with 50 real-valued entries.

Promising alternatives to nonlinear optimization can be found in the literature, which involve reducing the composite identification problem into two separate, linear identification problems, one for each function of Figure 17. This would be readily possible if the intermediate signal $v(n)$ was available from the measurements of the RFPA, but it is not. However, it can be approximated in one of two ways. One possible approach [28] involves expressing the intermediate signal as both a function G of $x(n)$, and a function H^{-1} of $y(n)$, where $x(n)$ and $y(n)$ are available data from the RFPA input and output, G is the function comprising the first block of the cascade in Figure 17, and H^{-1} is the *inverse* of the function comprising the second block. These simultaneous expressions involving $v(n)$ are then rearranged to yield a linear expression involving $x(n)$, $y(n)$, and the coefficients of G and H^{-1}, which can be solved to find $v(n)$. With $v(n)$ available, both the first and second blocks of the cascade can be identified separately with LSE, using the dataset pairs $x(n)$ and $v(n)$, and $v(n)$ and $y(n)$, respectively. Another method [41] involves constructing a forward model of the RFPA using the Volterra series, which can mimic the RFPA's transfer characteristics with a high degree of accuracy. This forward model is then fed with a sufficiently low power signal which is assumed to stimulate only the linear memory effects and not the dynamic nonlinearity; this is a reasonable assumption, since for values of $x(n-m) \ll 1$ dynamically

nonlinear terms of the form $x(n-m)\left|x(n-m)\right|^k$ become negligible. The 'small signal' output and input of the forward model can then be used to extract coefficients of an FIR filter using LSE; these coefficients are substituted into the first block of the cascade, so that the intermediate signal $v(n)$ can now be estimated. With $v(n)$ available, the second block can be identified. While the first method above has only been used for the identification of Wiener models, the latter approach has been applied to the proposed model with success [41]. However, both approaches involve repeated applications of LSE, which involves matrix inversions and the manipulation of large data vectors; hence, computationally friendlier identification schemes, such as RLS, should be investigated as future work.

3.4 Single-band Linearization Results

The performance of the proposed single-band FIR-APD model was tested by using it to linearize a wideband GaN broadband Doherty PA [42] driven to near-saturation at a peak output power of 43 dBm with communication signals of progressively increasing modulation bandwidth. The setup of the experimental testbed is shown in Figure 18.

Since this thesis concerns the development and testing of an analog predistorter formulation and not the analog integrated circuitry itself, the coefficient estimator and predistorter engine have been realized completely in MATLAB, and the non-idealities of a physical APD engine have been abstracted to allow a bare assessment of the predistorter formulation. An arbitrary waveform generator performs bits-to-RF conversion on the digital baseband data that is uploaded to it via MATLAB. The driver amplifier is used to adjust power of the RF signal so that the RFPA is driven close to peak power, which in is verified using the power meter. The attenuator brings the output of the RFPA to a level that is suitable for the digitizer, while the downconverter shifts the output to a low intermediate frequency, such that the digitizer can perform digital downconversion on the signal. The digitizer feeds this recovered baseband signal back to the MATLAB engine, which uses it to re-identify predistorter coefficients in each iteration.

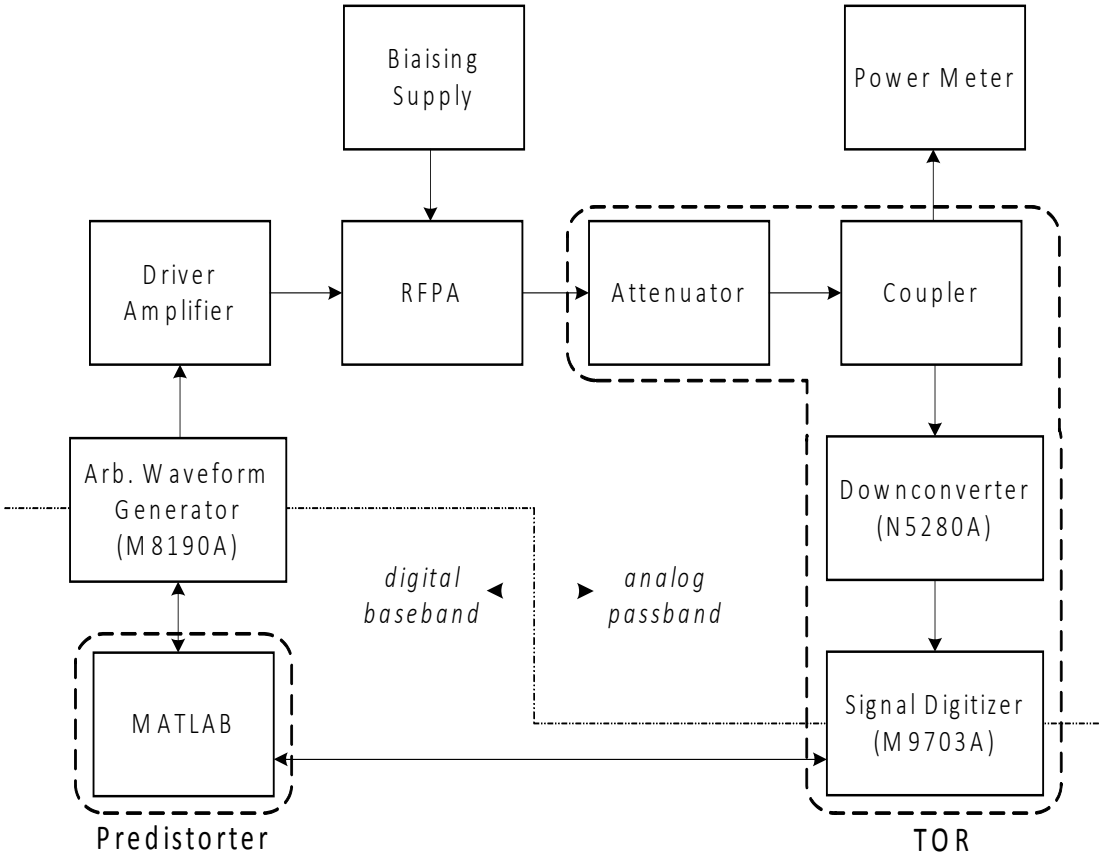

Figure 18: Experimental setup for the validation of single-band FIR-EMP

The following steps are performed at each iteration to identify the coefficients:

(i) Predistorter coefficients from the previous step are used to generate a sequence of 100,000 IQ points (corresponding to a 1ms slice of predistorted signal) that are uploaded to the AWG.

(ii) The AWG continuously transmits the predistorted sequence until the digitizer has captured a chunk of the RFPA output signal. The transmitted and captured signal are then used to identify the coefficients of the RFPA pre-inverse using least-squares (as per the process described in section 1). These coefficients replace the previous ones, and are used to generate the predistorted signal for the next iteration.

(iii) The above process usually needs to be repeated for 2-3 iterations until the predistorter performance becomes relatively steady from one iteration to the next – when this occurs, the final iteration records relevant measures of performance (such as EVM and ACPR) for RFPA operation with and without the predistorter. These are presented below.

The performance of the proposed FIR-EMP formulations is documented in Table 1 and contrasted against that of other single-band formulations; the linearized output spectra of the RFPA are shown in Figures 19-21 for three test signals of progressively higher modulation

bandwidth and PAPR (chosen to stimulate more non-ideal effects make linearization progressively more challenging): (i) a WCDMA '1001' signal (20 MHz BW) (ii) an intra-band non-contiguous carrier-aggregated WCDMA '111' and 15 MHz LTE signal (40 MHz total BW), and (iii) another intra-band non-contiguous carrier-aggregated WCDMA '1111' and 20 MHz LTE signal (80 MHz total BW).

Table 1: Performance comparison of single-band predistorter formulations

Signal	PAPR (dB)	BW (MHz)	Predistortion Scheme	NMSE (dB)	- ACPR-L (dB)	- ACPR-U (dB)	EVM (%)	# of Coeff.
WCDMA 4C-1001	7.11	20	No pred.	--	33.1	28.7	8.4	--
			Volterra DDR	-43.9	51.9	51.3	0.8	91
			MP	-40.1	50.7	51.4	2.3	21
			FIR-EMP (NL)	-47.0	50.3	49.8	0.9	22
WCDMA 3C-111 - LTE15	8.4	40	No pred.	--	36.1	29.8	6.4	--
			Volterra DDR	-42.8	48.9	48.1	1.0	91
			MP	-38.3	48.0	47.3	1.7	21
			FIR-EMP (NL)	-46.1	47.9	46.2	1.1	22
WCDMA 4C-1111 - LTE20	9.6	80	No pred.	--	38.2	29.0	7.1	--
			Volterra DDR	-39.5	46.6	45.0	1.4	91
			MP	-35.0	44.4	44.2	2.4	21
			FIR-EMP (NL)	-43.0	45.4	42.7	1.6	22

In all cases, the DDR Volterra, with its abundance of coefficients, represents the best achievable performance against which to compare the other formulations'. While all of the predistortion schemes suffer as the signal bandwidth increases, the modelling capability of the standard memory polynomial degrades rapidly compared to the Volterra series – this is evident from the increase in NMSE, and accompanying increase in out-of-band distortion (ACPR) and in-band distortion (EVM). The proposed FIR-EMP however, not only outperforms the MP using the same number of coefficients, but successfully competes with the DDR Volterra, using only a fraction of its coefficients (1/4th). These results provide evidence that the series of transformations that lead from the MP to the FIR-EMP retains essential modelling terms. Furthermore, only 1-2 dB of ACPR suppression is lost by using the FIR-EMP model instead

of the Volterra, suggesting that the FIR-EMP employs the bulk of essential terms needed to correct most of the nonlinearity and memory effects, and that introducing additional terms provides diminishing returns in terms of linearization performance. In light of the considerable design challenges and circuit non-idealities of an APD implementation, additional modelling terms of Volterra and Volterra-like formulations are perhaps best avoided.

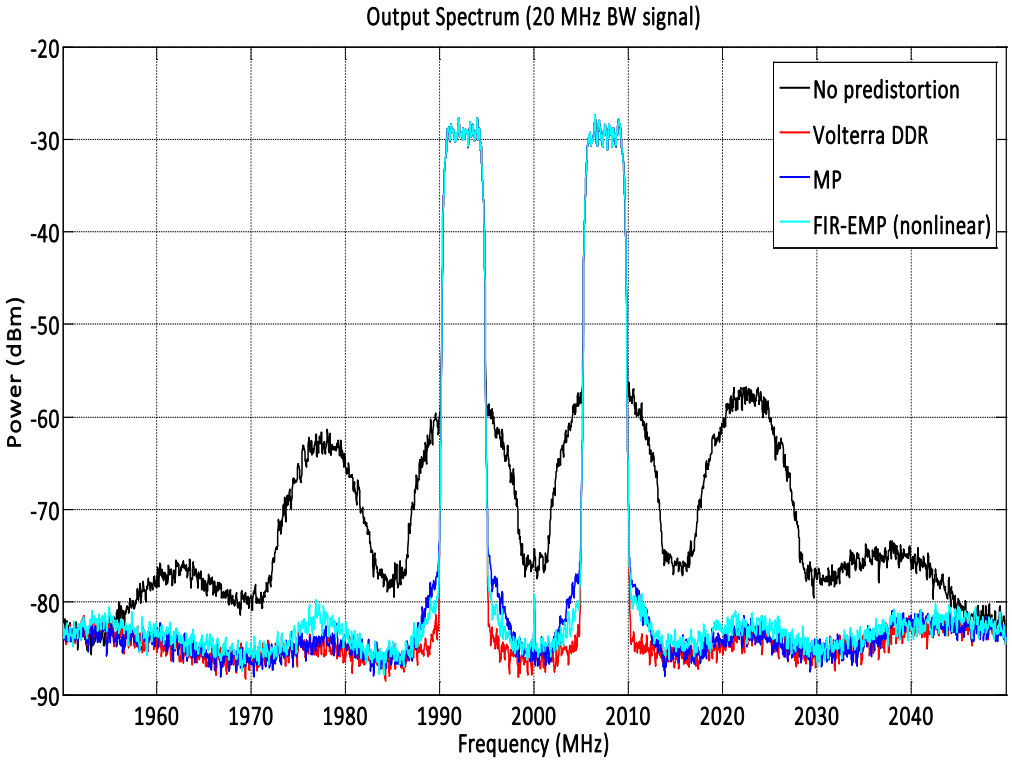

Figure 19: RFPA output spectrum for 20 MHz single-band signal

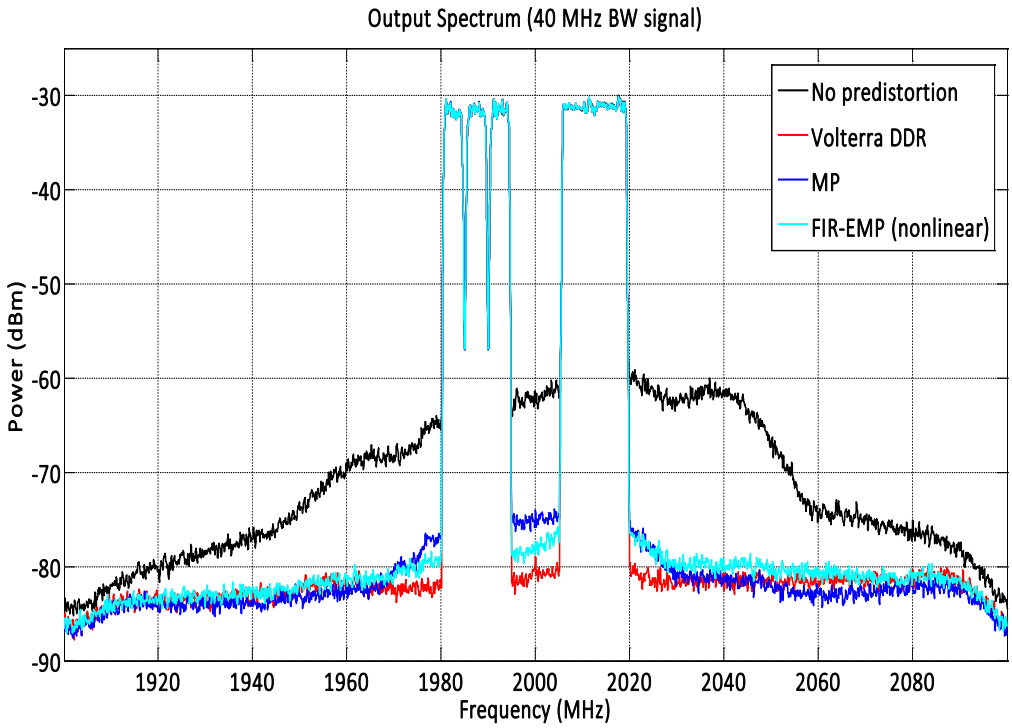

Figure 20: RFPA output spectrum for 40 MHz single-band signal

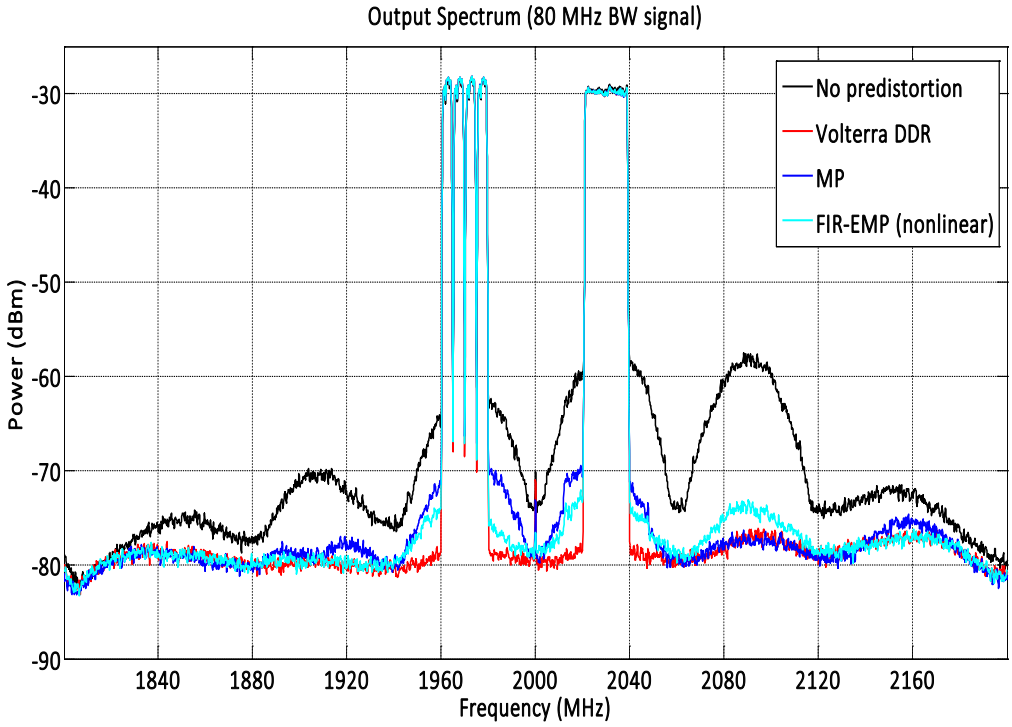

Figure 21: RFPA output spectrum for 80 MHz single-band signal

Chapter 4
Dual-band Analog Predistortion

This chapter is organized as follows. Section 4.1 provides the motivation for extending the single-band FIR-EMP formulation to dual-band, and Section 4.2 describes the extension and proposes a dual-band equivalent of the APD architecture from Chapter 3. In Section 4.3, the problem of coefficient identification in a dual-band context is discussed, and the single-band identification scheme previously proposed is extended to dual-band identification. Section 4.4 presents measurement results for the proposed formulation using the same RFPA as before, but under dual-band excitation.

4.1 Motivation for Dual-band Transmission

Wireless transmitters capable of concurrent multi-band transmission, i.e. the simultaneous transmission of multiple modulated signals at multiple, potentially widely separated frequencies, have been an area of great interest. Concurrent multi-band transmission can enable multi-functionality in wireless devices, such as the ability to interface with the global positioning system, transmit data over wireless local area networks, and make voice calls, all at the same time. Simultaneous transmission over multiple frequency bands also increases the effective transmission bandwidth and hence, data throughput of wireless systems. Additionally, it can improve the frequency diversity of the system, reducing its susceptibility to path loss and interference which are especially problematic for high frequency transmission in dense, urban environments. Given the success of single-band linearization using FIR-EMP, its extension to concurrent dual-band transmission scenarios would be desirable.

4.2 Extension to Dual-band

While dual-band transmitters initially employed separate RFPAs designed to achieve optimum efficiency in each of their respective bands, advances in design techniques have enabled the realization of single RFPAs to concurrently amplify the signals in both bands, which significantly reduces the cost and size of the transmitter. In such cases, where a nonlinear RFPA is fed with a dual-band signal that is the sum of the signals of the individual bands, a new class of distortion products appear at the output. In addition to the harmonic and intermodulation distortion present in single-band scenarios, dual-band amplification gives rise to cross-modulation distortion products, which are the result of interactions between frequency components of *both* bands. To account for these distortion products, dual-band versions of previously mentioned functions such as the Volterra series and the memory polynomial have been proposed in the literature, which include nonlinear cross-terms [43, 44]. The dual-band FIR-EMP (2D-FIR-EMP) function must be similarly formulated, and the procedure for deriving it is similar to the single-band case. The dual-band equivalent of the memory polynomial, shown in Eq. 4.1, is turned into an envelope-based formulation by disposing of terms that are problematic for an APD implementation. The resulting 2D-EMP is then augmented with FIR filters (one for each band) to provide the 2D-FIR-EMP formulation (Eq.

4.2). The block diagram of the 2D-FIR-EMP predistorter is shown in Figure 22, and should be contrasted with Figure 16. Duplication of hardware (such as the envelope detector, delay element, and vector multiplier) is necessary in the transition from single-band to dual-band.

$$
\begin{aligned}
x_{2D-DPD(1)}(t) &= \sum_{m=0}^{M}\sum_{k=0}^{N-1}\sum_{j=0}^{k} a_{m,k}^{(1)} x_{(1)}(t-m\tau)\left|x_{(1)}(t-m\tau)\right|^{k-j}\left|x_{(2)}(t-m\tau)\right|^{j} \\
x_{2D-DPD(2)}(t) &= \sum_{m=0}^{M}\sum_{k=0}^{N-1}\sum_{j=0}^{k} a_{m,k}^{(2)} x_{(2)}(t-m\tau)\left|x_{(2)}(t-m\tau)\right|^{k-j}\left|x_{(1)}(t-m\tau)\right|^{j}
\end{aligned}
\tag{4.1}
$$

$$
\begin{aligned}
x_{2D-FIR-EMP(1)}(t) &= \sum_{m=0}^{M}\sum_{k=0}^{N-1}\sum_{j=0}^{k} b_{m,k}^{(1)} x_{FIR(1)}(t)\left|x_{FIR(1)}(t-m\tau)\right|^{k-j}\left|x_{FIR(2)}(t-m\tau)\right|^{j} \\
x_{2D-FIR-EMP(2)}(t) &= \sum_{m=0}^{M}\sum_{k=0}^{N-1}\sum_{j=0}^{k} b_{m,k}^{(2)} x_{FIR(2)}(t)\left|x_{FIR(1)}(t-m\tau)\right|^{k-j}\left|x_{FIR(2)}(t-m\tau)\right|^{j} \\
x_{FIR(1)}(n) &= \sum_{v=0}^{V} a_{v}^{(1)} x_{(1)}(n-v) \\
x_{FIR(2)}(n) &= \sum_{v=0}^{V} a_{v}^{(2)} x_{(2)}(n-v) \qquad (t=nT)
\end{aligned}
\tag{4.2}
$$

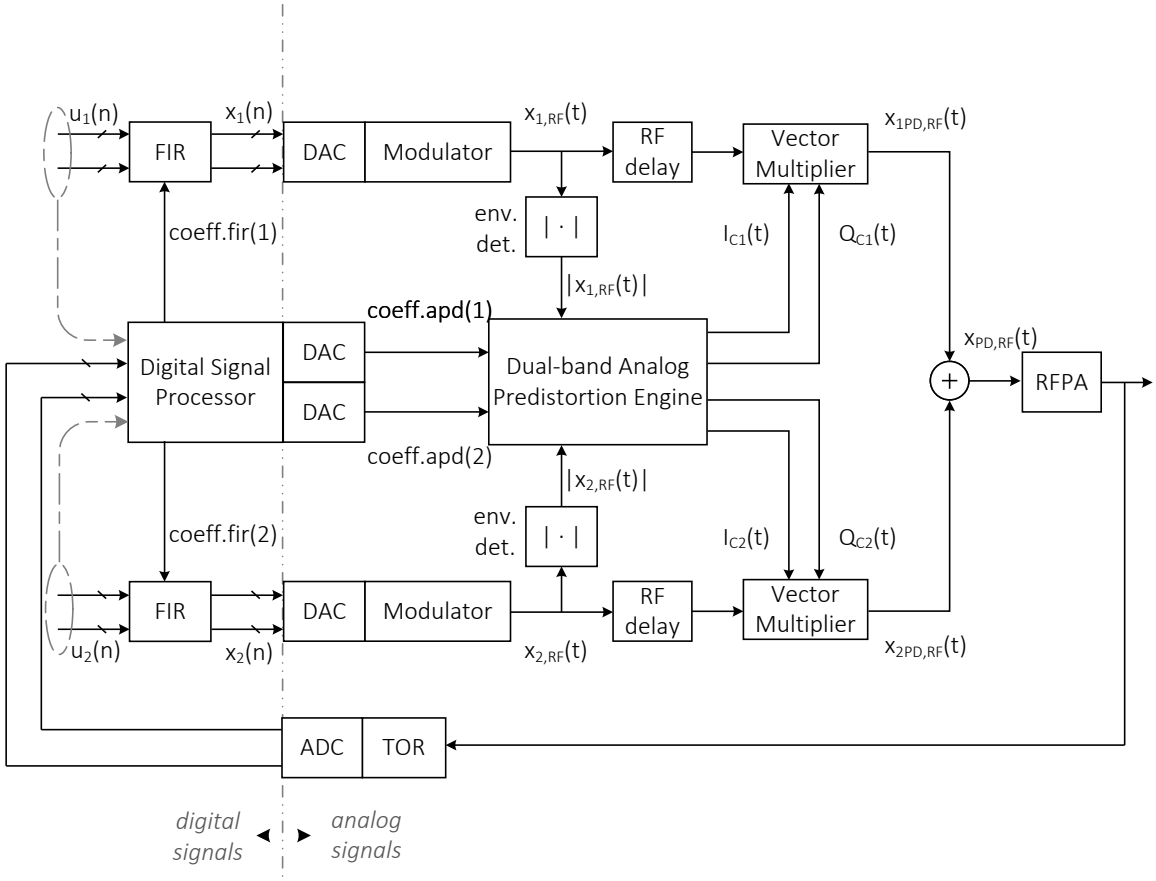

Figure 22: Block diagram of the FIR-APD system for a dual-band predistorter

An analog block implementation of the 2D-FIR-EMP is possible, but the visualization of it is fairly onerous; hence, a hardware block diagram at a higher level of abstraction is presented in Figure 23. The figure shows arrays of analog delay blocks generating the delayed versions of $|x_1(t)|$ and $|x_2(t)|$; for each delay branch, these envelope signals are fed into a separate nonlinear basis generator – a network of multipliers that self-multiply and cross-multiply the envelopes of the two bands – to generate the required bases of the 2D-FIR-EMP. These nonlinear signals are then forwarded to the coefficient multiplier blocks, where arrays of low power DACs and multipliers are used to scale the bases according to the real and imaginary parts of the complex coefficients. The correction signals from each delay branch are then summed to generate the overall I and Q correction signals, one for each band.

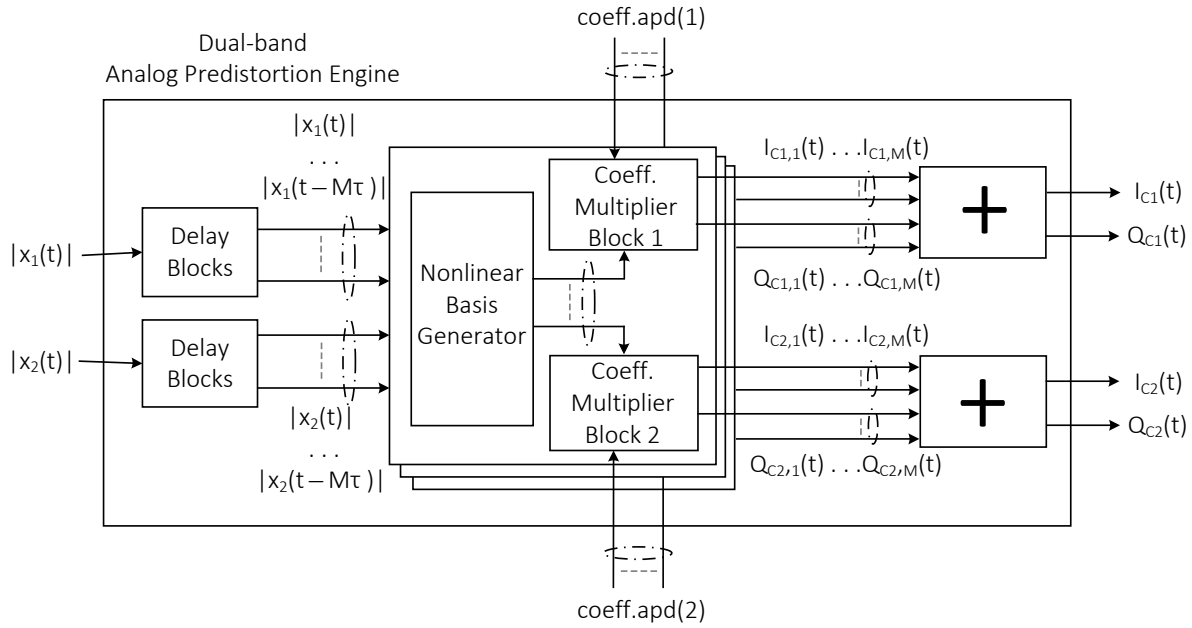

Figure 23: Block diagram of a dual-band analog predistorter engine

A quick inspection of Eq. 4.2 reveals that the dual-band formulation would have roughly quadruple the number of bases (and coefficients) compared to the single-band case (see Eq. 3.4). The first doubling occurs simply because each of the bands in a dual-band scenario require their own predistorter formulation; another doubling occurs within the predistorter formulations in each band – in addition to generating the intra-band distortion that is present in single-band scenarios, nonlinear RFPAs amplifying dual-band signals generate cross-band distortions that need to be linearized using additional bases. As a result of this cross-band distortion, the predistorter formulations for band-1 and band-2 need to be nonlinear functions of the input from *both* bands. Because of the added complexity this will introduce to the analog engine, in terms of the required number of multipliers and adders that need to be realized, a pruning of the 2D-FIR-EMP model is necessary. An assumption of 'fading memory' can be made on the non-ideal nature of the RFPA, which allows the memory depth of the formulation to be progressively reduced as the nonlinearity order increases. This

assumption has been proposed in prior literature [45, 46], in which its usage was justified on rather empirical grounds. However, progressive truncation of the memory depth is a reasonable proposition, considering the physical origins of these non-ideal effects. Section 1 discussed how interactions between intermodulation distortion currents and the drain impedance of the RFPA gave rise to dynamically nonlinear behavior. Typically, distortion products that are more recent have greater impact on the instantaneous drain modulation that affects the RFPA output; furthermore, higher order distortion products are generally smaller than lower order ones and have less impact on the drain modulation. As a result, the coefficients a_{mk} corresponding to dynamic nonlinear bases of the form $x(n-m)\left|x(n-m)\right|^k$ typically become smaller as m and k become large, to the point where those bases can simply be neglected without compromising linearization performance. Thus, a pruned formulation of the 2D-FIR-EMP is presented in Eq. 4.3 in which the memory depth $M(k)$ varies according to the nonlinearity order k; this pruned formulation will be used in the rest of the thesis.

$$
\begin{aligned}
x_{2D-FIR-EMP(1)}(t) &= \sum_{m=0}^{M(k)}\sum_{k=0}^{N-1}\sum_{j=0}^{k} b_{m,k}^{(1)} x_{FIR(1)}(t)\left|x_{FIR(1)}(t-m\tau)\right|^{k-j}\left|x_{FIR(2)}(t-m\tau)\right|^{j} \\
x_{2D-FIR-EMP(2)}(t) &= \sum_{m=0}^{M(k)}\sum_{k=0}^{N-1}\sum_{j=0}^{k} b_{m,k}^{(2)} x_{FIR(2)}(t)\left|x_{FIR(1)}(t-m\tau)\right|^{k-j}\left|x_{FIR(2)}(t-m\tau)\right|^{j} \\
M(k) &= M-\left(\frac{k}{2}\right) \qquad (4.3) \\
x_{FIR(1)}(n) &= \sum_{v=0}^{V} a_v^{(1)} x_{(1)}(n-v) \\
x_{FIR(2)}(n) &= \sum_{v=0}^{V} a_v^{(2)} x_{(2)}(n-v) \qquad (t=nT)
\end{aligned}
$$

4.3 Dual-band Predistorter Model Identification

The 2D-FIR-EMP model is nonlinear in its coefficients, like its single-band counterpart. Similar to what was done in section 3.2, the nonlinear optimizer will first be proposed to identify coefficients that can demonstrate the potential of the 2D-FIR-EMP formulation; decoupled least squares estimation will then be proposed as a less computationally intensive alternative. However, the cascaded nature of the proposed predistorter poses a challenge to coefficient identification that is specific to dual-band (and multi-band scenarios), which is discussed below.

From the dual-band EMP formulation, it is evident that each of the correction signals of band-1 (and band-2) is a nonlinear function of the signal envelopes of both bands. These signal envelopes, in turn, are linear, filtered versions of the undistorted signal, as per Eq. 4.2. As a result, the correction signal for band-1 is affected by the 2D-EMP coefficients of the first band $b_{m,k}^{(1)}$, and the FIR filter coefficients of both bands, $a_v^{(1)}$ and $a_v^{(2)}$. Likewise, the correction

signal for band-2 is affected by the 2D-EMP coefficients of the second band $b_{m,k}^{(2)}$, and the FIR filter coefficients of both bands, $a_v^{(2)}$ and $a_v^{(1)}$. This co-dependency can be problematic if one chooses to identify the optimal coefficients of the two bands independently – a tactic that is commonly employed in dual-band digital predistortion schemes where the respective formulations of the bands are decoupled from one another.

When using nonlinear optimization, the co-dependency issue can be circumvented by redefining the separate problems of finding optimal solutions in each band, to a single optimization problem. This is achieved by defining the coefficient vector to be optimized as c_{2D}, which is the concatenation of the coefficient vectors of the two bands, and the cost function $f_{obj,2D}(c_{2D})$ to be the average of the normalized mean square error between the measured and estimated input of the PA in the two bands (Eq. 4.5).

$$f_{obj,2D}(c_{2D}) = \frac{1}{2}\left(\frac{\sum_{i=0}^{L}\left|x_{1,i} - f_{2D-FIR-EMP}(c_{2D}^{(1)}, y_{1,i}, y_{2,i})\right|^2}{\sum_{i=0}^{L}\left|x_{1,i}\right|^2}\right) + \frac{1}{2}\left(\frac{\sum_{i=0}^{L}\left|x_{2,i} - f_{2D-FIR-EMP}(c_{2D}^{(2)}, y_{2,i}, y_{1,i})\right|^2}{\sum_{i=0}^{L}\left|x_{2,i}\right|^2}\right)$$

$$\begin{aligned} c_{2D} &= [c_{2D}^{(1)} \;\; c_{2D}^{(2)}]^T \\ c_{2D}^{(1)} &= [a_v^{(1)} \;\; b_{m,k}^{(1)}]^T \\ c_{2D}^{(2)} &= [a_v^{(2)} \;\; b_{m,k}^{(2)}]^T \end{aligned} \qquad (4.5)$$

Just as in Section 3.2, the Quasi-Newton approach was used to search for the optimal coefficients. However, the search space for the concatenated coefficient vector will be larger compared to its single-band counterpart, and will require considerably more iterations to converge on an optimal solution. The computationally less intensive alternative to this approach is to extend the decoupled linear identification approaches discussed in Section 3.2 to the dual-band case, which has been recommended for future work. The approach would remain the same, but it would need to be performed twice, once to find $c_{opt,2D}^{(1)}$ based on signals $x_1(t), y_1(t), y_2(t)$, and once to find $c_{opt,2D}^{(2)}$ based on signals $x_2(t), y_2(t), y_1(t)$. Whether this approach would be significantly impacted by the coefficient co-dependency issue, described above, is a matter of further investigation.

4.4 Dual-band Linearization Results

The performance of the proposed dual-band FIR-APD model was tested by using it to linearize the same broadband Doherty PA of Section 3.4, driven to near-saturation at a peak output power, using an inter-band carrier aggregated 15 MHz bandwidth WCDMA ‘101’ signal and a 15 MHz bandwidth LTE signal with 300 MHz carrier frequency separation, and a combined PAPR of 9.3 dB. The setup of the experimental testbed was very similar to that of the single-

band case, except that for the dual-band signal, the downconverter creates intermixing products that, once captured by the digitizer, interfere with the fidelity of the feedback signal. Thus, the downconverter and digitizer were replaced with a N9030A PXA signal analyzer.

The performance of the proposed FIR-EMP scheme is documented in Table 2 along with that of other dual-band linearization schemes that are extensions of the single-band memory polynomial and BBE Volterra schemes. The linearized output spectra of the RFPA for band-1 and band-2 are shown in Figures 24-25. The proposed 2D-FIR-EMP, when identified with nonlinear optimization, can successfully compete with 2D-DPD, even when it is pruned to have a fraction of the coefficients. These results provide indisputable evidence of essential modeling terms being preserved through the transformations that lead us from the 2D-DPD to the 2D-FIR-EMP.

Table 2: Performance comparison of dual-band predistorter formulations

Signal	BW (MHz)	Predistortion Scheme	NMSE (dB)	- ACPR-L (dB)	- ACPR-U (dB)	EVM (%)	# of Coeff
WCDMA 3C-101 (Band 1)	15	No pred.	--	35.8	35.1	4.7	--
		2D-BBE Volterra	-43.7	51.9	52.5	0.7	27
		2D-DPD	-43.5	52.2	51.8	0.7	45
		2D-FIR-EMP (NL)	-43.4	51.6	52.3	0.7	20
LTE15 (Band 2)	15	No pred.	--	35.5	34.5	4.6	--
		2D-BBE Volterra	-43.9	52.4	53.4	0.7	27
		2D-DPD	-44.3	51.8	53.0	0.7	45
		2D-FIR-EMP (NL)	-44.1	51.3	51.5	0.8	20

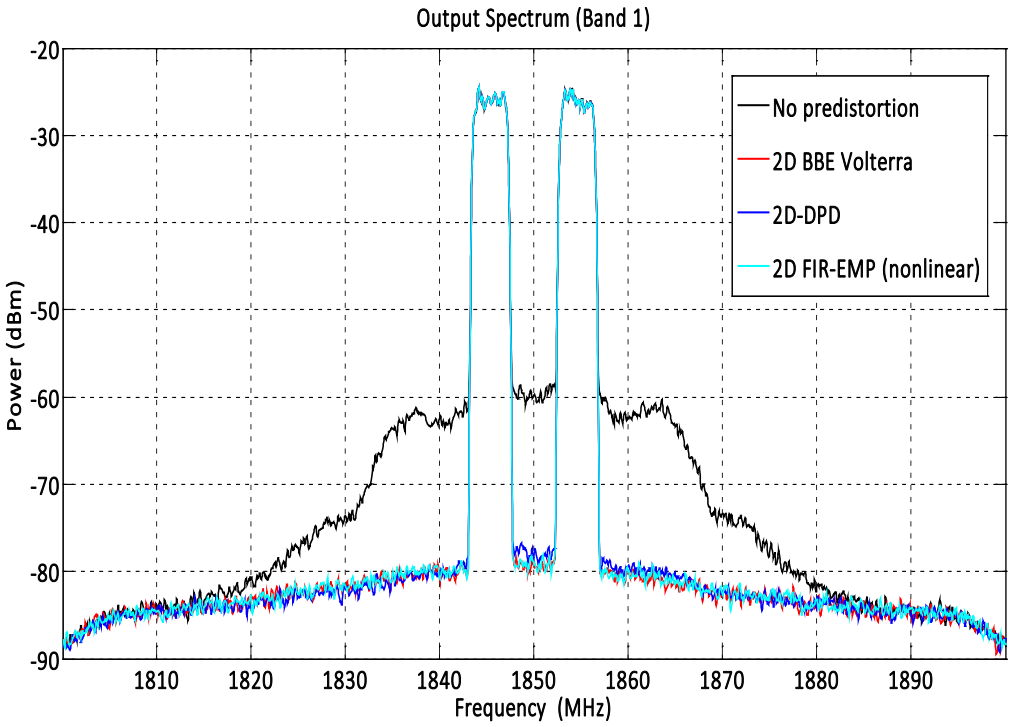

Figure 24: RFPA band-1 output spectrum for dual-band signal

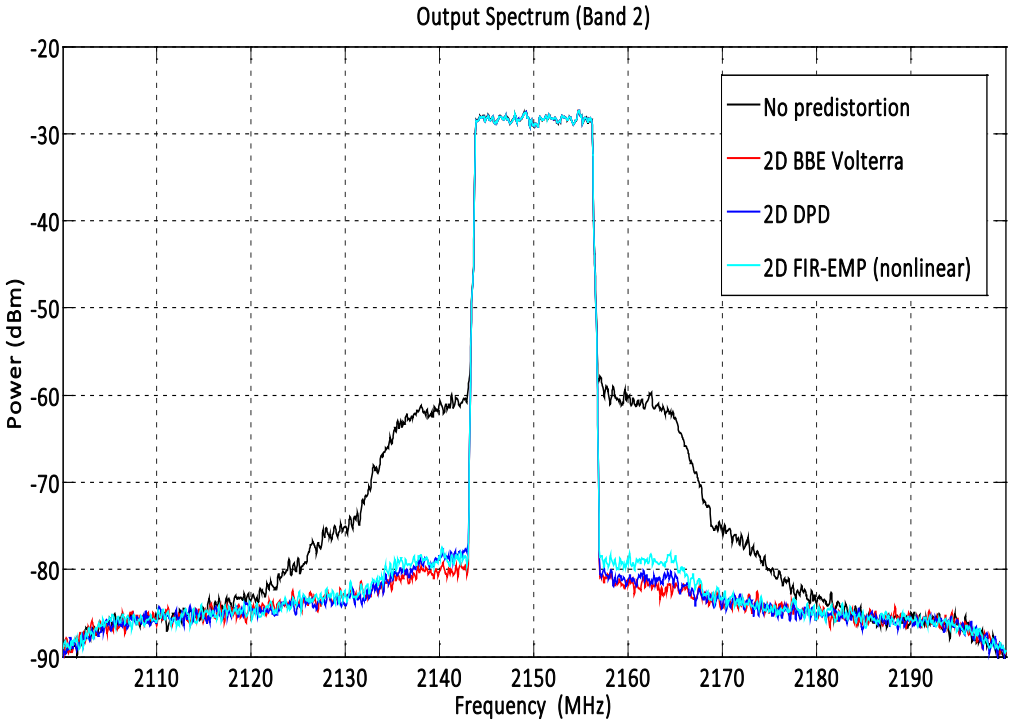

Figure 25: RFPA band-2 output spectrum for dual-band signal

Chapter 5
Conclusions and Future Work

5.1 Conclusions

As low-power small cells employing highly nonlinear power amplifiers are increasingly adopted into common usage, traditional RFPA linearization techniques such as digital predistortion need to be replaced with low-power, affordable alternatives such as analog predistortion. The objective of this thesis was to propose a predistorter formulation that is well-adapted to the APD architecture and the constraints it imposes, while performing competitively against well-established DPD formulations. Based on the research underlying this thesis, the following conclusions can be drawn:

Straightforward adaptations of DPD formulations to APD are precluded by hardware limitations in generating the required mathematical bases. The envelope memory polynomial is an obvious choice for APD schemes due to: (i) its low complexity, (ii) the fact that it can be expressed as a complex gain correction on the undistorted signal, and (iii) that fact that this gain is only a function of the undistorted signal magnitude (and not the phase).

However, the performance of the EMP suffers as modulation bandwidth of the communication signal increases, limiting its scope to narrow-band communication. The linearization failure was attributed to the loss of linear memory terms in the transition from MP to EMP, and this was confirmed by noting the recovery in performance when the EMP is augmented with a parallel FIR filter.

More importantly, the modelling benefit of the linear memory terms can be realized even when the FIR is cascaded before the EMP. The cascade structure allows the FIR filter to be realized trivially in digital baseband, while the EMP is realized in analog – thus allowing both the advantages of accurate digital signal processing and low-power analog circuitry to be leveraged for predistortion.

As demonstrated through measurement results for a GaN Doherty RFPA, the resulting FIR-EMP formulation can successfully predistort single-band signals of up to 80 MHz modulation bandwidth, even achieving better linearity than the memory polynomial model from which it was derived. However the cascade model is nonlinear with respect to the unknown coefficients and precludes the use of least squared methods. Nonlinear coefficient identification using the Quasi-newton approach was used to demonstrate the unhindered performance of the model.

The proposed FIR-EMP formulation and nonlinear coefficient estimation were extended to dual-band transmission. To stymy the growth in coefficients that arise in dual-band scenarios, a coefficient pruning based on the assumption of fading memory assumption was used. The pruned 2D-FIR-EMP formulation retained competitive performance with other dual-band extensions of well-known predistortion schemes such as the memory polynomial and baseband equivalent Volterra.

5.2 Future Work

It is evident that coefficient identification can be the limiting factor in realizing the benefits of the proposed analog predistortion formulation. As an alternative to nonlinear estimation, the use of alternating least squares should be investigated to improve the estimation accuracy. Also known as the co-ordinate block descent approach, this algorithm would assume initial coefficients for the FIR and EMP blocks, and update them in an alternating manner to reduce the combined mean-squared error of the model. To avoid the matrix inversions and manipulation of large data vectors associated with the pseudo-inverse based least squares, recursive least squares should be employed for the alternating estimation. Adopting a recursive approach would allow the use of a direct learning architecture, in which access to the input of the RFPA is no longer required. As acknowledged in Section 3.1, this is essential to realizing the power savings of the APD architecture.

Ultimately, the performance of the proposed FIR-EMP scheme should be validated by implementing the EMP formulation in an analog circuit, where physical non-idealities such as circuit noise, finite dynamic range, device mismatch, and time-delay misalignment can limit the performance. Finally, given the success of FIR-EMP in single and dual-band, its performance should be investigated for tri-band scenarios, similar to what has been done for the memory polynomial, and the Volterra series.

References

[1] H. Sarbishaei, Yushi Hu, B. Fehri and S. Boumaiza, "Concurrent dual-band envelope tracking power amplifier for carrier aggregated systems," in *Microwave Symposium (IMS), 2014 IEEE MTT-S International,* 2014, pp. 1-4.

[2] H. Golestaneh, F. A. Malekzadeh and S. Boumaiza, "An Extended-Bandwidth Three-Way Doherty Power Amplifier," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 61, pp. 3318-3328, 2013.

[3] Chi-Chia Hsieh and E. Strid, "A S-band high power feedback amplifier," in *Microwave Symposium Digest, 1977 IEEE MTT-S International,* 1977, pp. 182-184.

[4] R. G. Meyer, R. Eschenbach and W. M. Edgerley, "A wide-band feedforward amplifier," *Solid-State Circuits, IEEE Journal of,* vol. 9, pp. 422-428, 1974.

[5] M. Coers and W. Bosch, "DC to 6.5 GHz highly linear low-noise AlGaN/GaN traveling wave amplifier with diode predistortion," in *Microwave Symposium (IMS), 2014 IEEE MTT-S International,* 2014, pp. 1-4.

[6] Min-Gun Kim, Chung-Hwan Kim, H. Yu and Jaejin Lee, "An FET-level linearization method using a predistortion branch FET," *Microwave and Guided Wave Letters, IEEE,* vol. 9, pp. 233-235, 1999.

[7] Guang Yang, Haoyu Wang, Lingli Li and Falin Liu, "One-step model extraction method for direct learning digital predistortion," *Electronics Letters,* vol. 50, pp. 1148-1150, 2014.

[8] R. N. Braithwaite and S. Carichner, "An improved doherty amplifier using cascaded digital predistortion and digital gate voltage enhancement," in *Microwave Symposium Digest, 2009. MTT '09. IEEE MTT-S International,* 2009, pp. 1073-1076.

[9] L. Ding, G. T. Zhou, D. R. Morgan, Zhengxiang Ma, J. S. Kenney, Jaehyeong Kim and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *Communications, IEEE Transactions on,* vol. 52, pp. 159-165, 2004.

[10] O. Hammi, F. M. Ghannouchi and B. Vassilakis, "A Compact Envelope-Memory Polynomial for RF Transmitters Modeling With Application to Baseband and RF-Digital Predistortion," *Microwave and Wireless Components Letters, IEEE,* vol. 18, pp. 359-361, 2008.

[11] B. Fehri and S. Boumaiza, "Baseband Equivalent Volterra Series for Behavioral Modeling and Digital Predistortion of Power Amplifiers Driven With Wideband Carrier Aggregated Signals," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 62, pp. 2594-2603, 2014.

[12] F. Mkadem, M. Fares, S. Boumaiza and J. Wood, "Complexity-reduced Volterra series model for power amplifier digital predistortion," ***Analog Integr. Cir. Signal Proc.*****, vol. 79, pp. 331-343, 05/01, 2014.**

[13] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in ***Wireless Communication Systems. 2008. ISWCS '08. IEEE International Symposium on,*** **2008, pp. 309-313.**

[14] T. Gotthans, R. Marsalek, M. Pospisil, J. Blumenstein and G. Baudoin, "Experimental verification and comparison of polynomials and LUTs predistortion techniques," in ***Radioelektronika (RADIOELEKTRONIKA), 2015 25th International Conference,*** **2015, pp. 309-312.**

[15] Wangmyong Woo, M. D. Miller and J. S. Kenney, "A hybrid digital/RF envelope predistortion linearization system for power amplifiers," ***Microwave Theory and Techniques, IEEE Transactions on,*** **vol. 53, pp. 229-237, 2005.**

[16] R. N. Braithwaite, "Memory correction for a WCDMA amplifier using digital-controlled adaptive analog predistortion," in ***Radio and Wireless Symposium (RWS), 2010 IEEE,*** **2010, pp. 144-147.**

[17] F. Roger, "A 200mW 100MHz-to-4GHz 11th-order complex analog memory polynomial predistorter for wireless infrastructure RF amplifiers," in ***Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International,*** **2013, pp. 94-95.**

[18] S. C. Cripps, "Reduced conduction Angle—Waveform analysis," in ***RF Power Amplifiers for Wireless Communications,*** **2nd ed. ed.Anonymous Boston: Artech House, 2006, pp. 42-46.**

[19] (2013, October 9). ***Understanding Intermodulation Distortion Measurements.*** **Available: http://electronicdesign.com/communications/understanding-intermodulation-distortion-measurements.**

[20] J. Vuolevi, T. Rahkonen and J. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," in ***Radio and Wireless Conference, 2000. RAWCON 2000. 2000 IEEE,*** **2000, pp. 195-198.**

[21] S. C. Cripps, "PA memory effects," in ***RF Power Amplifiers for Wireless Communications,*** **2nd ed. ed.Anonymous Boston: Artech House, 2006, pp. 256-260.**

[22] A. E. Parker and J. G. Rathmell, "Bias and frequency dependence of FET characteristics," ***Microwave Theory and Techniques, IEEE Transactions on,*** **vol. 51, pp. 588-592, 2003.**

[23] T. R. Cunha, J. C. Pedro and E. G. Lima, "Low-pass equivalent feedback topology for power amplifier modeling," in *Microwave Symposium Digest, 2008 IEEE MTT-S International,* 2008, pp. 1445-1448.

[24] J. C. Pedro, N. B. Carvalho and P. M. Lavrador, "Modeling nonlinear behavior of band-pass memoryless and dynamic systems," in *Microwave Symposium Digest, 2003 IEEE MTT-S International,* 2003, pp. 2133-2136 vol.3.

[25] J. Tsimbinos and K. V. Lever, "Computational complexity of Volterra based nonlinear compensators," *Electronics Letters,* vol. 32, pp. 852-854, 1996.

[26] Anding Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned volterra series," *Microwave and Wireless Components Letters, IEEE,* vol. 14, pp. 563-565, 2004.

[27] Anding Zhu, J. C. Pedro and T. R. Cunha, "Pruning the Volterra Series for Behavioral Modeling of Power Amplifiers Using Physical Knowledge," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 55, pp. 813-821, 2007.

[28] P. L. Gilabert, G. Montoro and E. Bertran, "On the wiener and hammerstein models for power amplifier predistortion," in *Microwave Conference Proceedings, 2005. APMC 2005. Asia-Pacific Conference Proceedings,* 2005, pp. 4 pp.

[29] Taijun Liu, S. Boumaiza, A. B. Sesay and F. M. Ghannouchi, "Dynamic nonlinear behavior characterization for wideband RF transmitters using augmented hammerstein models," in *Microwave Conference, 2006. APMC 2006. Asia-Pacific,* 2006, pp. 967-970.

[30] H. Enzinger, K. Freiberger and C. Vogel, "Analysis of even-order terms in memoryless and quasi-memoryless polynomial baseband models," in *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on,* 2015, pp. 1714-1717.

[31] Maxim Integrated. (2017, July 28). *RF PREDISTORTION (RFPD) VS. DIGITAL PREDISTORTION (DPD)*. Available: https://www.maximintegrated.com/en/products/comms/wireless-rf/rf-predistorters/predistortion-linearization/rfpal-technology/rfpd-vs-dpd.html.

[32] Analog Devices. AD9743/AD9745/AD9746/AD9747: Dual, 10-/12-/14-/16-bit 250 MSPS, digital-to-analog converters data sheet (rev. B). [Online]. Available: http://www.analog.com/media/en/technical-documentation/data-sheets/AD9743_9745_9746_9747.pdf.

[33] Guangxiang Yuan, Xiang Zhang, Wenbo Wang and Yang Yang, "Carrier aggregation for LTE-advanced mobile communication systems," *Communications Magazine, IEEE,* vol. 48, pp. 88-93, 2010.

[34] T. Rosowski and et al, "Preliminary spectrum scenarios and justification for WRC agenda item for 5G bands above 6 GHz," Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS)-II, Tech. Rep. R3.1, Oct. 2015.

[35] F. Mkadem and S. Boumaiza, "Nonlinear system behavioral modeling using reduced transmitter observation receiver bandwidth," in *Microwave Measurement Conference (ARFTG), 2014 83rd ARFTG,* 2014, pp. 1-3.

[36] A. Eghbali, H. Johansson, O. Gustafsson and S. J. Savory, "Optimal Least-Squares FIR Digital Filters for Compensation of Chromatic Dispersion in Digital Coherent Optical Receivers," *Lightwave Technology, Journal of,* vol. 32, pp. 1449-1456, 2014.

[37] S. J. Savory, "Digital Coherent Optical Receivers: Algorithms and Subsystems," *Selected Topics in Quantum Electronics, IEEE Journal of,* vol. 16, pp. 1164-1179, 2010.

[38] D. F. Shanno, "Conditioning of Quasi-Newton Methods for Function Minimization," *Mathematics of Computation,* vol. 24, pp. 647–656, 1970.

[39] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *The Computer Journal,* vol. 7, pp. 308-313, January 01, 1965.

[40] A. R. Conn, N. I. M. Gould and P. L. Toint, *Trust-region methods.* Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[41] 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," *Microwave Theory and Techniques, IEEE Transactions on,* vol. PP, pp. 1-9, 2015.

[42] M. N. Abadi, "Extended Bandwidth Doherty Power Amplifier for Carrier Aggregated Signals," *UWSpace,* 2014.

[43] B. Fehri and S. Boumaiza, "Baseband Equivalent Volterra Series for Digital Predistortion of Dual-Band Power Amplifiers," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 62, pp. 700-714, 2014.

[44] S. A. Bassam, M. Helaoui and F. M. Ghannouchi, "2-D Digital Predistortion (2-D-DPD) Architecture for Concurrent Dual-Band Transmitters," *Microwave Theory and Techniques, IEEE Transactions on,* vol. 59, pp. 2547-2553, 2011.

[45] J. Staudinger, J. -. Nanan and J. Wood, "Memory fading volterra series model for high power infrastructure amplifiers," in *Radio and Wireless Symposium (RWS), 2010 IEEE,* 2010, pp. 184-187.

[46] J. Staudinger, "DDR volterra series behavioral model with fading memory and dynamics for high power infrastructure amplifiers," in *Power Amplifiers for Wireless and Radio Applications (PAWR), 2011 IEEE Topical Conference on,* 2011, pp. 61-64.