RF Power Amplifier Modeling and Linearisation for Improved Training and Dimension Reduction Techniques

Méabh Loughman

Supervised by Dr. John Dooley and Prof. Ronan Farrell

Maynooth University
National University of Ireland Maynooth

A thesis submitted in partial fulfilment of the requirements

for Doctor of Philosophy

Department of Electronic Engineering

Maynooth University

28/02/22

Declaration

I hereby certify that this thesis, which I now submit for assessment on the programme of study leading to the award of PhD has not been submitted, in whole or part, to this or any other University for any degree and is, except where otherwise stated the original work of the author.

Signature: Méabh Loughman

Dedication

To my family and friends, of all species.

Contents

List of Figures

List of Tables

Abstract

The energy used to power cellular communication networks has grown immensely in recent times due to the increased human reliance on the wireless transmission of data. This has led to increased energy consumption of cellular networks, intensified by the advent of the Fifth Generation (5G) communication standard. In order to improve the energy efficiency of current and future cellular communication generations, the reasons why and how to reduce the power consumption without affecting performance is an important area of research.

Power Amplifiers (PAs) are responsible for a considerable percentage of the power inefficiencies in Radio Frequency (RF) transmitters. There is a trade-off in the topology of PA architectures between linearity and efficiency. However, to maintain both high efficiency and linearity external linearisation procedures can be implemented digitally. Digital Predistortion (DPD) has been demonstrated as a suitable linearisation solution of PAs that enables the retention of high efficiency and signal linearity. However as the carrier frequencies and bandwidths of the power amplifiers increase along with the introduction of multi-antenna base stations, new research and implementation obstacles emerge regarding efficiency, estimation and computational complexity of traditional DPD methodologies.

This thesis contributes to the advancement of PA models, DPD function estimation process in terms of accuracy and robustness, and finally dimension reduction of specific DPD functions. These novel developments are cultivated with the aforementioned energy targets of 5G transmission and reception taken into consideration.

Four novel contributions are presented in this thesis. The four contributions are related to publications listed in the Introduction of this thesis. The focal point of work presented is optimisation of techniques for PA modelling and linearisation.

The first novel contribution was enhancing a specific learning adaption, Recursive Least Squares, with improved robustness. An elegant early stopping criterion was established to minimize both time taken to train model coefficients and model accuracy.

The next novel contribution was a new strategy for DPD to combat the presence of multi-collinearity, and further reduce computational cost. The next contribution is related to the calculation of the limits to which modern communication signals can be down-sampled before use in calculating DPD coefficients , regardless of the DPD function used.

The final contribution is that of a novel methodology for calculating a common set of coefficients suitable for behavioural model and DPD coefficient estimation for any modulated signal. Native features of the testbench were used and the structures implemented employed hardware resources commonly found in Field Programmable Gate Arrays (FPGAs).

Chapter 1

Introduction

Telecommunication infrastructure has an increasing economic and environmental impact on the world [1]. The technological development required directly translates to demands placed on mobile networks [2]. Fifth Generation (5G) cellular network communications have promised key improvements, such as reduced latency and higher data throughput, compared to that of previous cellular networks as seen in Figure 1.1 [3]. These improvements require a rapid advancement in performance.

Preceding generations of cellular networks focused on design objectives regarding the optimisation of network characteristics. 3G consisted of multiple core cellular networks. 4G focused on an integrated, global network based on an open system approach. 4G was designed to replace the multiple core cellular networks with a single worldwide cellular core network standard based on Internet Protocol (IP) for control, video, packet data and telephony [4]. Power consumption has increased for 5G due to a number of factors, for instance multiple input multiple output transciever systems are becoming more prevalent, increasing the architecture or bill of materials needed. The design of and implementation of 5G and future cellular networks must now further improve network energy efficiency to combat the increased power consumption.

The necessity of the evolution of wireless communications, with continuously in-

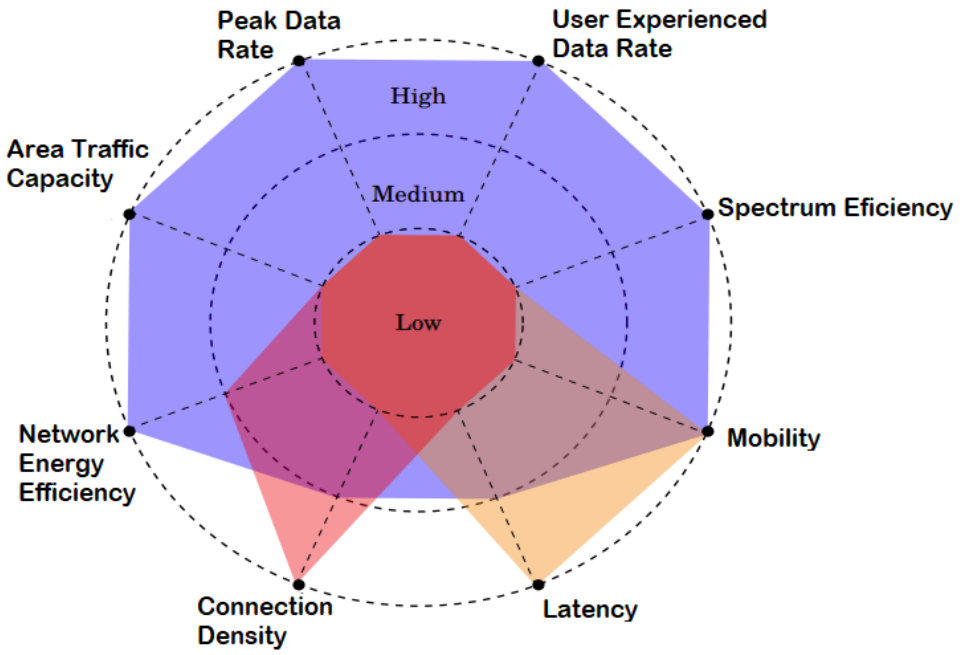

Figure 1.1: Increased importance of demands for 5G networks

creasing demands for higher data rates and capacity, is increasing the complexity of Radio Frequency (RF) transmitters. To meet the demands of future wireless communication systems, extensive research is being conducted to develop more energy efficient, re-configurable radio transmitters capable of supporting multiple radio access technologies and operates in a diverse range of frequency bands.

A significant component in the configuration of such complex radio transmitters is the power amplifier (PA). PA's are important components of signal transmission chains as they are responsible for increasing the power of the communication signals to power levels that are suitable for transmission. PA's are typically one of the last components in the radio transmitter chain- and therefore transmit and amplify a large amount of RF power. The efficiency at which PAs convert direct current (DC) power into RF power plays a critical role in the overall power consumption of a wireless communication system.

To avoid interference with other wireless communications and reduce the energy consumption in RF transmitters, PAs are required to behave linearly and efficiently.

This linearity-efficiency trade off is a historical issue. PA's can be 'backed off',i.e. operate at a lower average power output, to maintain linearity or alternate methods must be implemented to elicit the desired linear behaviour.

1.1 Motivation

It is clear that developing methodologies to increase power efficiency, and therefore linearisation techniques, is not only desirable from a communication operator's perspective but a requirement from an environmental stand point. The motivation of this thesis is the linearisation of nonlinear RF amplifiers. Digital Pre-Distortion (DPD) is a method by which input signals to a PA are weighted in order to produce a linear transmission signal as an output of the PA as depicted by Figure 1.2.

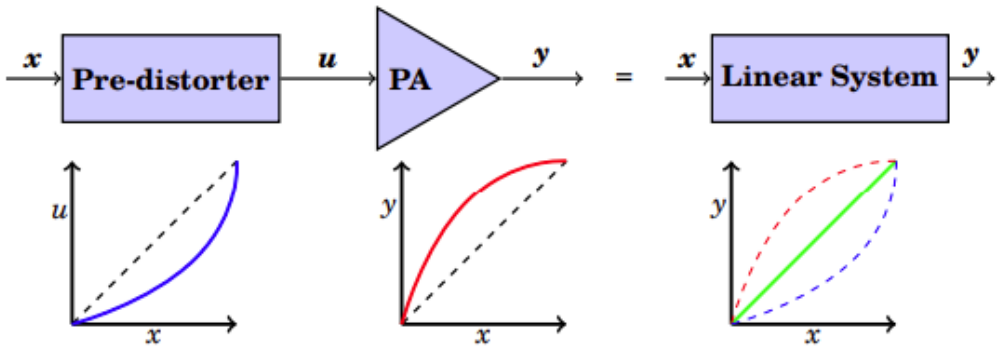

Figure 1.2: Principle of DPD

Linearity during transmission is a legal requirement that must comply with set operating frequencies, bandwidths and output power requirements. DPD is an attractive solution for linearisation due to its performance and ease of integration. DPD is capable of compensating for the distortion that is inherent when transmitting from high efficiency PAs. DPD research gives rise to:

1. Enable high efficiency transmitter architecture to comply with legalisation
2. Find techniques to compensate nonlinear distortion that said efficient transmitters introduce

1.2 Thesis outline

This thesis presents four new DPD techniques. Subsequent chapters of this thesis are organised as follows:

- Chapter 2 gives a high level description of a typical transceiver chain and relevant wireless communications background theory to this thesis.

- Chapter 3 presents existing behavioural modeling and pre-distorter methods. This thesis focuses on methods to enhance the use of polynomial based behavioural modeling and DPD.

- Chapter 4 discusses the contribution of this work regarding an early stopping technique designed to identify the onset of instability when training PA models using the Recursive Least Squares adaption technique.

- Chapter 5 proposes a novel methodology to combat the presence of multi-collinearity of basis functions used for DPD. The novel technique allows for local DPD coefficient updates, reducing the computational complexity when performing DPD. Updating locally, or in a partitioned manner, allows for a more computationally efficient DPD coefficient update as only selected DPD coefficients are updated iteratively on an as needed basis.

- Chapter 6 introduces a novel metric to establish the level of pre-processing of an input signal to a DPD/Behavioural model to ensure its output signal retains sufficient similarity, enabling accurate extraction of coefficients to be used in DPD or behavioural modeling.

- Chapter 7 provides a DPD methodology utilising a novel look-up table indexing technique, based on AMAM curves.

- Chapter 8 consists of concluding remarks and a discussion on possible future work.

1.3 Research contributions

The contribution of this thesis is to improve polynomial behavioural modeling and DPD techniques. The following list of chapters discuss in further detail each chapter in which novelty is introduced.

1.3.1 Chapter 4 - Ensuring Stability for DPD training

Contributions in Chapter 4 address the implementation of an early stopping technique, to avoid instability, of the Recursive Least Squares technique when training memory polynomial based behavioural models. The contribution of this work relates to multiple demands placed on 5G networks, as seen in Figure 1.1. Without the proposed elegant method to eradicate instability of recursive least squares other more computationally complex methods to ensure stability, when used, would suffer from poor data accuracy and be inefficient in terms of energy used.

- M. Loughman, R.Farrell, and J.Dooley , "Early stopping criteria for adaptive training of dynamic nonlinear behavioural models", in 2019 30th Irish Signals and Systems Conference (ISSC), Maynooth,2019, pp 1-5

- M. Loughman, S. Barton, R.Farrell, and J.Dooley, "Early Stopping Criterion for Recursive Least Squares Training of Behavioural Models", Wireless Personal Communications, 2022

1.3.2 Chapter 5 - DPD Function Dimension Reduction

The contributions in Chapter 5 develop a novel DPD function for the purposes of increasing computational efficiency of updating coefficients locally, enabling matrix inversions to be more compact and achieve high performance. The metrics benefitting from this application, as in Figure 1.1, are that of latency, spectrum efficiency and network energy efficiency.

- M. Loughman, R.Farrell, and J.Dooley , "Acceleration of Digital Pre-Distortion

Training Using Selective Partitioning", The 2022 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR 2022)

1.3.3 Chapter 6 - Critical Subsampling Rate for High Frequency PA modeling

Contributions presented in Chapter 6 support ongoing research investigating the challenge of the output path of a PA requiring a sample frequency that is a multiple of the original signal bandwidth when conducting behavioural models and DPD. A metric that determines the most advantageous pre-processing method, in terms of NMSE, is discussed. The contribution of this work relates to multiple demands placed on 5G networks, as seen in Figure 1.1, such as reduced latency, should this methodology be adopted at base station.

- M. Loughman, R.Farrell, and J.Dooley , "A Metric for training signal dimension reduction for high frequency power amplifier modeling and DPD", In preparation

1.3.4 Chapter 7 - Efficiency in FPGA Implementation of DPD

The contribution contained within Chapter 7 demonstrate a novel look-up table indexing technique to help determine the optimal DPD coefficients needed for predistortion. The contributions, as depicted by Figure 1.1, are minimised latency, spectrum efficiency and Mobility.

- H. Zhaoyang, M.Loughman, Y.Jiang, R. Mushini, M. Leeser and J.Dooley , "Multi-standard Digital Pre-Distortion Look Up Tables", 17th EAI International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM 2021)

Chapter 2

Wireless Communications Systems

Wireless communication systems consist of both active and passive components. The configuration of components can be tailored for specific use cases, but there are some components that are common in most architectures, as shown in Figure 2.1. An analogue communication subsystem which is responsible for the transmission and reception of radio signals and a digital baseband system responsible for the digital signal processing operations applied to the information signals.

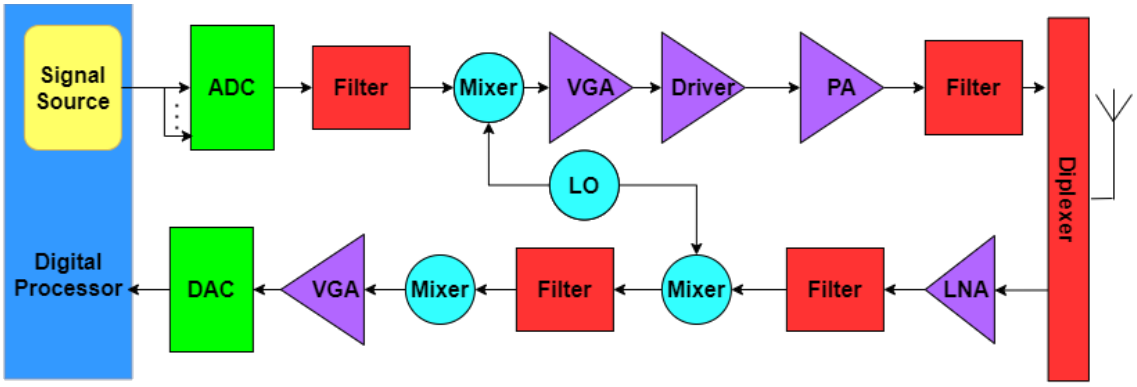

Figure 2.1: Components of modern radio

All radios have analogue hardware such as:

- Antennas and Matching Networks
- Filters
- Amplifiers
- Mixers

- Analogue to Digital Converters
- Digital to Analogue Converters

Each of the above elements have specific uses. Antennas and matching networks are used to ensure the maximum usable power is transferred to the output of the transceiver system. Filters exclude unwanted signal components, for example the bandpass filters eliminate components centred on harmonic frequencies. Two potential sources of harmonic frequencies are: the local oscillator (LO) and secondly nonlinear response of amplifiers. The driver amplifier is an optional amplifier used to pre-amplify an input signal into a PA if the ADC/DAC cannot deliver the power needed to excite the PA. The PA increases the power level of the RF signal to a suitable level for transmission. The efficiency of an amplifier is the ratio between the output RF power over the DC input power, given as a percentage as seen in (2.1).

$$\text{Efficiency} = \frac{\mathrm{P_{RF}}}{\mathrm{P_{DC}}} \times 100 \tag{2.1}$$

Mixers allow for up and down conversion of the RF signal to/from an intermediate frequency (IF) to be transmitted/received. ADCs and DACs allow for conversions between the digital and analogue domain. N-bit ADC outputs a N-bit binary number proportional to the analogue input voltage. The input range can be centred at zero volts or a nominal DC offset. The input is then divided into 2^N discrete values. Uniform quantisation, refers to the discrete values being evenly spaced and the discrete spacing can be described by equation (2.2).

$$\Delta = \frac{V_{fs}}{2^N} \tag{2.2}$$

Where V_{fs} refers to the voltage within the ADC range, otherwise the voltages are clipped. Dynamic range is the range of the input signal power over which the ADC performs adequately. The maximum input level is known as 'full scale'. Signal level

at the ADC input must be kept near full scale to avail of the maximum dynamic range.

2.1 Non ideal behaviour in wireless communications

Devices under test have a desired behavior, such as inductance (effects by a component that stores energy in a magnetic field), multiplication, or linear amplification. All devices have an operating region in which they exhibit this behavior, and other regions in which they do not. Noise and parasitic reactance are problematic when evaluating devices under test. The associated limits must be considered when using or designing any kind of radio system.

There are many analogue hardware undesired effects that can be categorised into distortion or disturbance. Distortion : The characteristics of the signal are affected. There is not distortion if there is not a signal. Disturbance : Undesired effects added to the signal. They exist if there is or there is not signal.

ADCs do not provide any gain or loss to a transciever chain but are affected by quantisation noise and thermal noise. The SNR at the output of a N-bit ADC is given by (2.3).

$$\mathrm{SNR_{ADC}(dB)} = 6.02\mathrm{N} + 1.76 \tag{2.3}$$

The higher the number of bits, the smaller the quantization error and therefore the quantization noise. DACs perform the inverse function of ADCs. DAC resolution is also given by N. Resolution refers to the smallest increment of output that the DAC can replicate as in (2.4).

$$\mathrm{DAC_{res}} = \frac{\mathrm{V_{fs}}}{2^{\mathrm{N}} - 1} \tag{2.4}$$

where V_{fs} refers to full scale voltage.

Receiver sensitivity is the minimum signal power needed at the receiver input to provide adequate Signal to Noise Ratio (SNR) to carry out the data demodulation. The required SNR depends on the communications standard and the modulation scheme.

In order for a component to be linear, the output behaves the same as the input, but with a different amplitude or phase (time delay), in other words a linear component is an active device which linearly transforms an input signal, x, into an output signal, y such that $y = Gx$. G is the gain experienced. No hardware devices are perfectly linear, but they may have a linear region in which the behaviour is approximately linear. Active devices behave according to their device architectures. For example, Bipolar transistors have an exponential characteristic and long channel MOSFETs have a square-law characteristic. Composite devices may have more complicated characteristics [5].

Approximates are utilised when mathematically describing these complicated characteristic functions, typically with a Taylor Series expansion as seen in (2.5).

$$V_o = a_1 V_{in} + a_2 V_{in}^2 + a_3 V_{in}^3 + ... \tag{2.5}$$

2.2 Non ideal behaviour of Power Amplifiers

Power consumption is a key factor in wireless sensor nodes and mobile terminals. PAs dominate power consumption in high power transmitters. Even in low-power transmitters, the PA may still rival the digital circuitry in power consumption. The primary tradeoff in PA design is efficiency versus linearity. Most efficient amplifiers are nonlinear, thus methods to elicit linear behavior while maintaining high efficiency is an area of extensive research.

Ideally, a PA linearly transforms an input signal, x, into an output signal, y, such

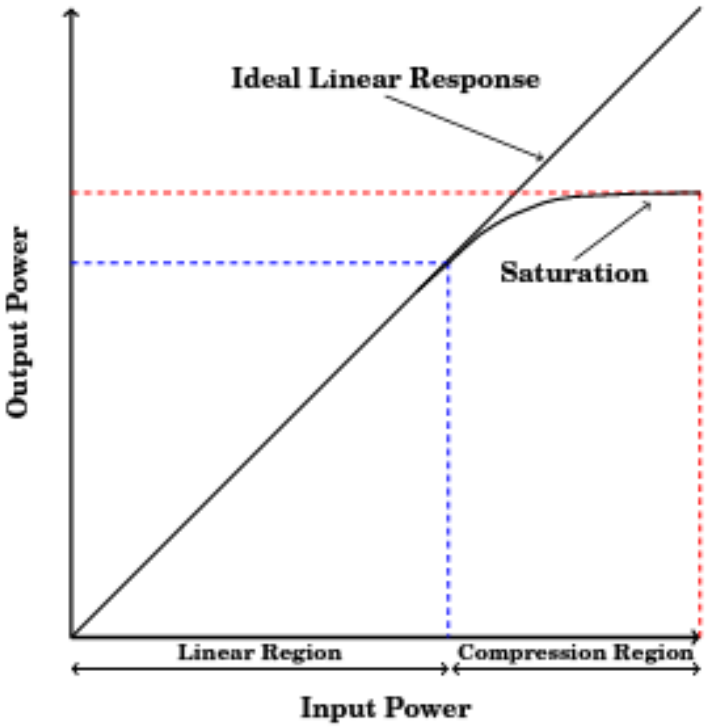

Figure 2.2: PA operation

that $y = Gx$, where G is the gain experienced. PAs, in reality, elicit nonlinear behaviour. The linearity versus efficiency is a historical trade-off.

Efficiency of a component, device or system, η, can be defined as the effective output power P_y divided by its input power P_x, given by equation (2.6).

$$\eta[\%] = \frac{P_{RF}}{P_{DC}} \tag{2.6}$$

An 'ideal' PA would transfer all DC power to the PA output signal. As PAs behave nonlinearly the highest efficiency is commonly achieved by exciting the PA close to its saturation region, where the PA will elicit nonlinear behaviour.

PAs maintain linear operation in a low efficiency operating region, i.e. very little of the DC power is converted to output RF signal power.

2.2.1 Nonlinearities of power amplifiers

2.2.2 Signal clipping

The modulation used in modern communication systems, such as 5G-NR 256 QAM, enables a high throughput of data. The modulation schemes modulate phase and amplitude of the signals to be sent. The input signal to the PA, will have a specific Peak-to-Average Power Ratio (PAPR), as seen in Figure 2.6. The transmitted signal may have high power peaks, causing further backing off to be necessary to combat distortion of the output, as shown in Figure 2.3.

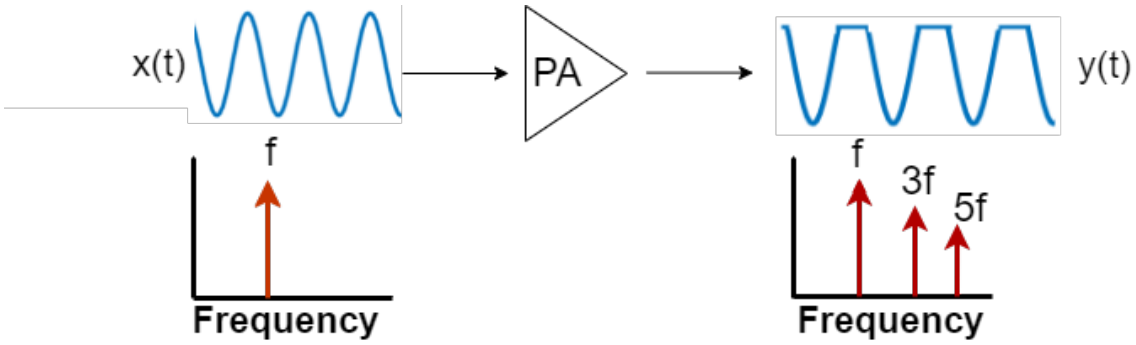

Figure 2.3: Illustration of signal clipping and presence of harmonic distortion

Different signal modulation schemes exhibit noticeable signal distortion in the time domain. As stated the PAPR depends on the modulation scheme used. Should the amplitude of a signal increase significantly, i.e. beyond tolerable power levels, that are dictated by the transistor architecture, the PA may be driven into saturation as seen in Figure 2.2. The previously described behaviour may result in an abrupt change from the linear region into the saturated region, resulting in a sharp discontinuity to the output signal, commonly referred to as clipping. Should this occur, information contained in the signal will be lost and may result in frequency spreading. An illustration of a signal being clipped can be seen in Figure 2.3.

2.2.3 Harmonic distortion

Harmonic distortion is caused by additional tones being generated in the frequency spectrum. The tones typically arise from voltage and current variations due to

deviations in frequencies. Harmonic distortion degrades the linearity of the PA, and can be described mathematically from equation (2.5), as equation (2.7).

$$y(t) = a_0 + \frac{a_2 X_{in}^2}{2} + (a_1 X_{in} + \frac{3a_3 X_{in}^3}{4}) \cos(\omega t) + \frac{a_2 X_{in}^2}{2} \cos(2\omega t) + \frac{a_3 X_{in}^3}{4} \cos 3\omega t \quad (2.7)$$

Additional tones are generated by passing a single tone, $Xcos(\omega t)$), through the non-linear polynomial function describing the PA. It can be seen from equation (2.7) that harmonic distortion may be described as the power, or amplitude squared, A^2, at a multiple of the fundamental frequency(A_{nf}^2) compared with the power at the fundamental frequency (A_f^2), as defined in equation (2.8).

$$\mathrm{HD_{nf}} = \frac{\mathrm{A_{nf}^2}}{\mathrm{A_f^2}} \quad (2.8)$$

Total Harmonic Distortion (THD) refers to the sum of all of the harmonic distortions present in the output signal, described by equation (2.9).

$$\mathrm{THD} = \frac{\sqrt{\sum_{\mathrm{n=2}}^{\mathrm{N}} \mathrm{A_{nf}^2}}}{\mathrm{A_f^2}} \quad (2.9)$$

2.2.4 Intermodulation distortion

Transmission of two (or more) frequencies, nominally ω_1 and ω_2, through a PA results in intermodulation distortion. The output signal of the PA will include the originally transmitted signal and images of the transmitted signal at multiple frequencies of $\omega_1 + \omega_2$ and $\omega_1 - \omega_2$, as illustrated by Figure 2.4.

Even ordered intermodulation products will occur out of band. Odd ordered intermodulation products will not only occur out of band but also close to the originally transmitted signal on the frequency spectrum. Should these odd order intermodulation products contain sufficient power, interference with the transmitted signal

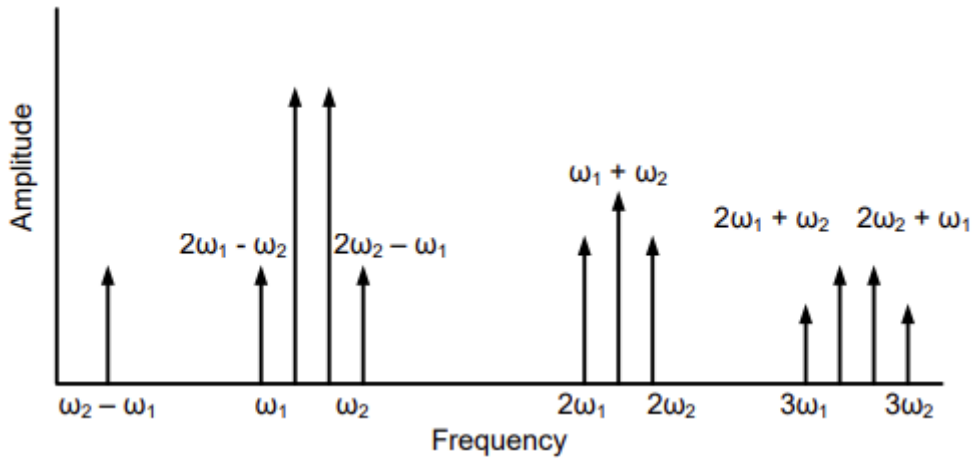

Figure 2.4: Illustration of intermodulation distortion

will occur, as well as with the frequencies/other users adjacent to it. The effects of strong intermodulation products will culminate in spectral defects, i.e. the transmission requirements of a given signal will not be maintained. This process, if not addressed, has the potential to produce bit errors in the transmission signal and will affect the transmitter in terms of efficiency with DC power being converted to redundant RF signals.

The effect of nonlinearity on the PA output signal can be visualised in both the frequency domain and time domain in Figure 2.5.

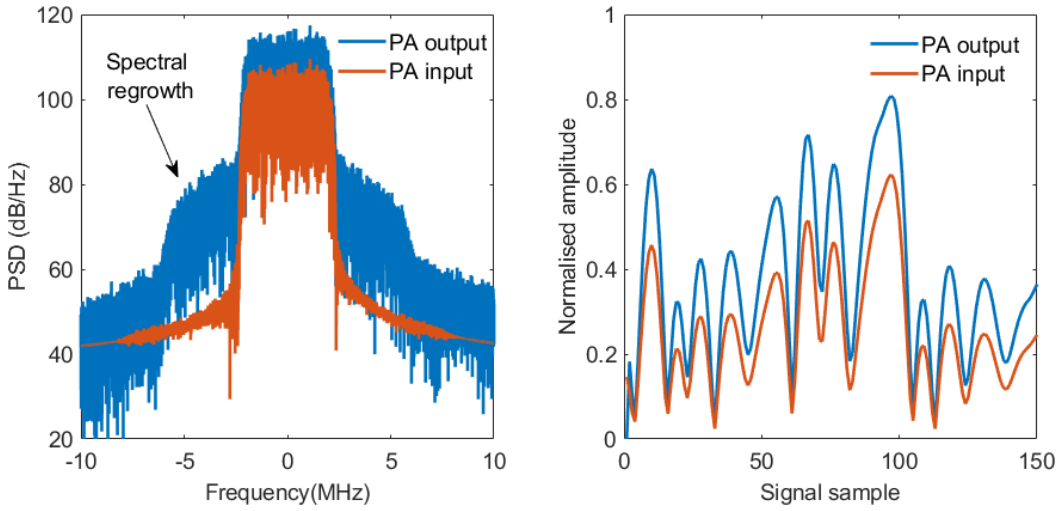

Figure 2.5: Experimental PA input and output signal, showing spectral regrowth and nonlinearity in the frequency domain and time domain. These signals were collected by transmitting a 5MHz WCDMA signal from MATLAB using an FMCOMMS3 through 10W RFHIC PA

2.2.5 Memory effects

Memory effects are deviations in the PAs output due to the causal relationship between current and preceding signal values. Memory effects can be categorized into two classifications, based on their period duration.

1. Long term memory effects
2. Short term memory effects

Long term memory effects are variations caused by temperature and aging equipment. Long term memory cause variations that are independent of signal bandwidth. Short term memory effects are typically caused by energy stored in the PA, such as in the drain supply, gate bias supply and the frequency response of the PA. The duration of short term memory effects are typically less than a few micro seconds and are directly related to signal bandwidth. The bandwidth relationship can be concluded by the passband of the PA and the charge/discharge time of electronic components within the PA. Short term memory effects can be mitigated against by using polynomials with memory capabilities to encapsulate PA output behaviour, as seen in equation (2.10). Long term memory effects may be lessened using methods such as assuming linear dependence of the parameters of a conventional model to a long term memory parameter, as in [6].

$$y(n) = \sum_{m=0}^{M} f_m(x(n-m)) \tag{2.10}$$

where n refers to the current input signal sample, M is the memory order and m refers to the memory delay.

2.3 Signal modulation

During transmission, the information is mapped into one of the properties of a sine wave; amplitude, frequency or phase. During reception of a signal at the receiver,

the sine wave is remapped and the mapped information is extracted. The sine wave is known as the carrier. The frequency of the carrier signal is the carrier frequency f_c and represents the frequency in the radio spectrum where the transmitted signal is centred. Radio frequency signals can be transmitted through mediums such as wire, air, fibre optic, water, space etc.

2.3.1 Orthogonality of signals

Two signals, $s_1(t)$ and $s_2(t)$, are orthogonal should their inner product be equal to zero, as seen in equation (2.11).

$$\langle s_1(t), s_2(t) \rangle = \int_a^b s_1^*(t) s_2(t) dt \tag{2.11}$$

Sinusoidal signals are commonly used in communications due to the ease of manipulation of orthognol signals. Sinusoidal signals at different frequencies are orthogonal if their frequencies are a multiple of a common reference frequency. A sine wave and a cosine wave with identical frequency are orthogonal, as seen in equation (2.12).

$$\langle \sin(2\pi f_c(t)), \cos(2\pi f_c(t)) \rangle = \int_0^T \sin(2\pi f_c(t)) \cos(2\pi f_c(t)) dt = 0 \tag{2.12}$$

2.3.2 Modulation

The information signal can be transmitted in the analogue or digital domain. In the analogue domain, the input signal directly modifies the amplitude, phase, or frequency of the carrier signal. Therefore, the information signal must be also analogue. One form of modulation is Amplitude Modulation (AM), where the input signal varies the amplitude of the carrier signal. AM is given by equation (2.13) and done by multiplying the information signal, $I(t)$, and the carrier signal, $\sin(2\pi f_c t)$.

$$s(t) = I(t) \sin(2\pi f_c t) \tag{2.13}$$

Quadrature Amplitude Modulation (QAM) superimposes two AM signals at the same frequency, but with the carrier 90 degrees out of phase. QAM is beneficial in the way it allows the frequency spectrum to be shared by two signals of equal bandwidth because the carriers are orthogonal due to the phase change of the carrier and can be seen in equation (2.14).

$$s(t) = I_1(t)\sin(2\pi f_c t) + I_2(t)\sin(2\pi f_c t + \frac{\pi}{2}) \tag{2.14}$$

Frequency Modulation (FM) varies the frequency of a carrier waves. Frequency and phase modulation can be seen in equations (2.15) and (2.16) respectively.

$$s(t) = A\cos((2\pi f_c(t) + k.I(t))) \tag{2.15}$$

$$s(t) = A\cos(2\pi f_c(t) + k\int_{-\infty}^{t} I(\tau)d\tau \tag{2.16}$$

PM signals are more resilient to nonlinearities in RF systems. Demodulation of PM signals is achieved by passing a signal through a bandpass filter designed to preserve only the fundamental frequency, and using a phase-locked loop (PLL), found in FPGAs, to extract the instantaneous frequency of the signal.

Digital modulation is also referred to as "symbol mapping" as the number of bits at the input are mapped onto a particular symbol given at the output. A digital information signal is expressed in bits and transmitted into the digital modulator at a particular bit rate f_b. A single bit (or more) are grouped to be mapped into a single symbol (M-bits per symbol). The output symbol rate is equal to the input bit rate divided by the number of bits per symbol. The pulse width filter (eq: Gaussian, Raised Co-sine and Sinc) is responsible for limiting the bandwidth of the signals.

Symbols can be represented as a point in two dimensional space on a constellation diagram. The X and Y co-ordinates represent the amplitude of an in-phase (cosine)

and quadrature (sine) component. A symbol represents one, or more, bits, contingent on the number of symbols. Co-ordinates of the X and Y on the constellation diagram is representative of the arrangements and movements of symbols undergoing diverse modulations at carrier frequency. There are a number of digital modulation schemes used throughout this thesis, Orthogonal Frequency Division Multiplexing (OFDM), Coded Division Multiple Access (CDMA), Amplitude Phase Shift Keying (APSK) and, Quadrature Amplitude Modulation (QAM).

QAM is modulated using a combination of AM and PM - the phase and amplitude are altered in tandem and an example can be seen in Figure 2.7. QAM allows for a high order of modulation which facilitates higher data transfer from the information signal. A trade-off is that high Signal to Noise Ratio (SNR) is necessary in order to mitigate against symbol errors.

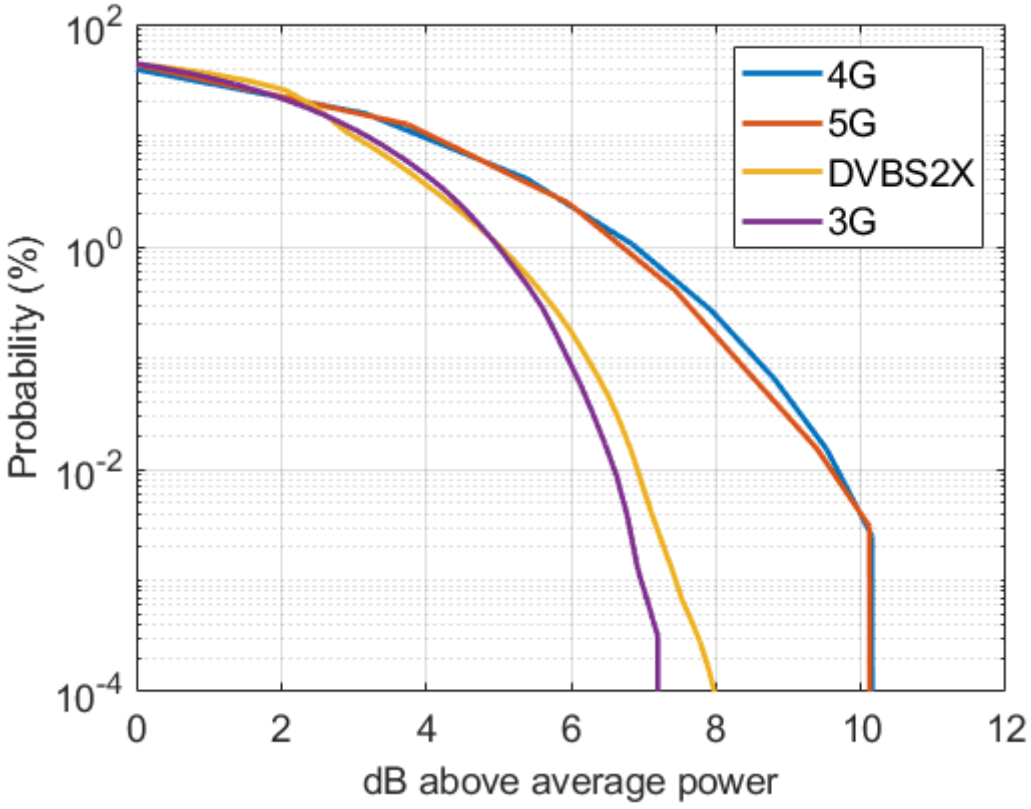

Figure 2.6: CCD vs PAPR. PAPR of multiple signal modulation schemes

WCDMA is implemented by spread spectrum techniques. Spread spectrum techniques multiply slower data by a high-speed code, the rate of which is called the

Table 2.1: Comparison of mobile communications signal standards at native bandwidths

Radio access technique	**Frequency range**	**Peak data rate**	**Modulation Schemes**	**PAPR (dB)**
3G	1.8- 2.5 GHz	2Mbps	WCDMA	6-8
4G	2 - 8 GHz	1Gbps	OFDM	6-12
5G	$>$ 6 GHZ 25-39 GHz	>1Gbps	CP-OFDM	10-12
DVB	950 - 2150 MHz	58.8 Mbps	APSK	6 -8

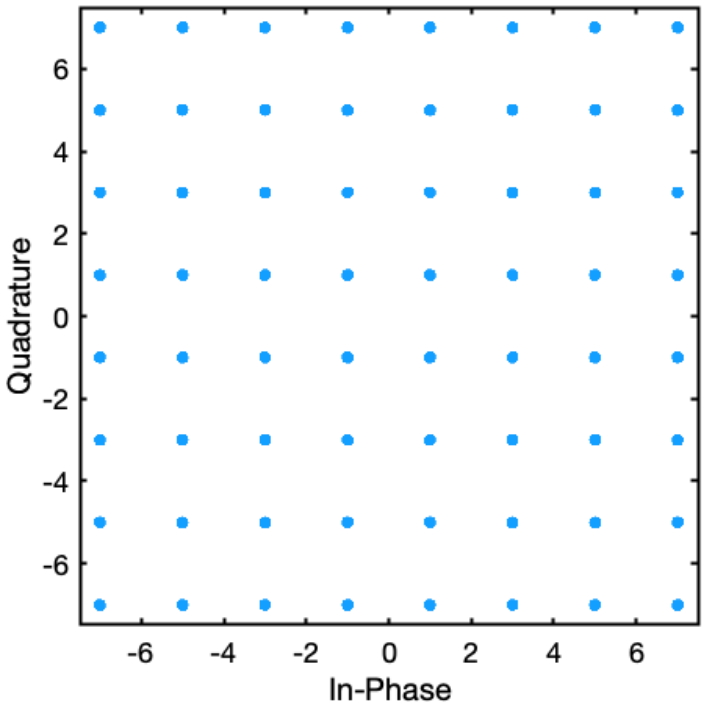

Figure 2.7: Example of 64QAM constellation diagram

chip rate, known as spreading sequence. The use of spread spectrum techniques allows for an expansion of bandwidth by the ratio of data rate to code rate (referred to as spreading factor) and multiple user access of the frequency channel. To implement spread spectrum digital modulation a symbol is multiplied by a spreading sequence or code associated with a particular user. The codes of two different users are orthogonal to each other, ensuring no interference between users.

OFDM is a multiple access modulation technique that allows for spectrum overlap between signals at multiple frequencies (as seen in equation (4.11)) resulting in sig-

nals that are uniquely recoverable. OFDM allocates the frequency band by dividing it into multiple smaller bands - allowing for users to use the channel constantly. Guard bands are typically used between the channels to avoid adjacent channel interference. The OFDM multiple access modulation technique is achieved by dividing a given high-bit-rate data stream into numerous parallel lower bit-rate streams and modulating each stream on separate carriers, called subcarriers. Subcarriers are chosen with the purpose that they are all orthogonal to one another. The first subcarrier is set to a frequency that it has an integer number of cycles in a symbol period. Subcarrier spacing between adjacent subcarriers is then set to BSC $= \frac{\mathrm{B}}{(\mathrm{L}-1)}$, where B is the nominal bandwidth and L is the number of subcarriers. This ensures the equivalent tones are orthogonal to one another over the symbol period. The advantage of OFDM is it's ability to combat intermodulation distortion.

APSK constellation diagrams are constructed such that the symbol points are configured in concentric rings of constant amplitude, as seen in Figure 2.8. An advantage of APSK is that it is a modulation scheme that can be structured to pre-distort itself. Altering the distance between the concentric rings prior to transmission enables pre-distortion of the signal in a way that counteracts the effects of transmission distortion [7].

There are legal requirements for transmission of wireless communications, however there are multiple standards in place due to specific applications working better under certain conditions. The standards examined in this thesis are described in the following sections.

3G - Universal Mobile Telecommunications System

Wideband Code Division Multiple Access (WCDMA) is used in the digital signal standard of 3G networks, otherwise known as Universal Mobile Telecommunications System (UMTS), and is standardized by the International Telecommunication Union used for mobile telecommunications networks that comply with the Interna-

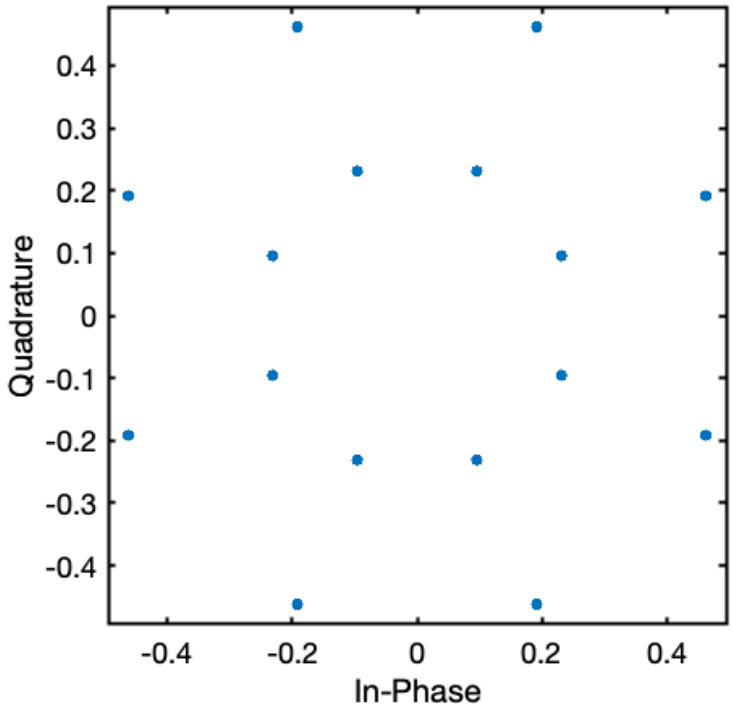

Figure 2.8: Example of APSK constellation diagram

tional Mobile Telecommunications-2000, specified by 3GPP [8]. WCDMA is a spread spectrum multiplexing communications scheme that allows individual orthogonal spreading codes to overlay multiple user data over the same 5MHz channel, using the following mapping techniques QPSK, 16 QAM and 64 QAM.

4G-Long Term Evolution

Orthogonal frequency-division multiplexing (OFDM), is a digital signal modulation scheme used in 4G networks, otherwise known as Long Term Evolution (LTE). LTE offers increased bandwidth of 20MHz per channel. OFDM is comprised of multiple sub carriers. Each sub carrier is capable of implementing separate modulation schemes. The resilience introduced by the ability to use disparate modulation schemes allows for flexibility compared to WCDMA, OFDM has a higher PAPR [9]. As discussed, increased bandwidth and high PAPR contribute to the nonlinear behaviour of a PA.

5G-New Radio

5G - New Radio(5G NR), is a radio access technology that also uses OFDM.

5G NR allows for simultaneous signal transmission and reception, and uses wider bandwidth channels. Both time and frequency domain modulation techniques allow for more efficient use of the frequency spectrum and includes guard bands to prevent propagation loss [10]. Carrier spacing is flexible in 5G NR, and allows for subcarrier modulation to be QPSK, 16QAM, 64QAM or 256QAM.PAPR reduction techniques must be implemented for 5G NR signals as, with OFDM, the signals have typically high PAPR [11]. 5G NR combats intersymbol interference (ISI) with the help of a cyclic prefix.

Digital Video Broadcasting

The Digital Video Broadcasting (DVB) standard is an application of DVB to satellite communications. In 2014, DVB-S2X was standardised to surpass the prior state of the art of DVB, DVB-S2. The main design improvement is that the available frequency spectrum is used more efficiently, supporting higher data rates using 32APSK, 64APSK or 256APSK modulation schemes [12].

2.3.3 Chapter summary

In summary, the main difference between the aforementioned radio access techniques or signal standards are the modulation schemes used and operating frequencies, as seen in Table 2.1. Different modulation schemes produce different PAPR. There are multiple sources of distortion when transmitting a signal through a PA. The contributions within this thesis offer methodologies such that the research solutions have been focused on 5G radio access techniques, with some research solutions applicable to all four standards introduced above, while combating the distortions introduced by signal transmission and reception through a PA.

Chapter 3

Power Amplifier Behavioural Modelling and Digitial Predisortion

PAs exhibit nonlinear behaviour, causing signal distortion in transmission. A signal sent through a PA will experience distortion at different power levels, dependent on what PA architecture is used. Ideally, PAs should operate with a high power output and illicit linear behaviour. This chapter will discuss elements critical to behavioural modeling and DPD, focusing on the PA in Figure 3.1.

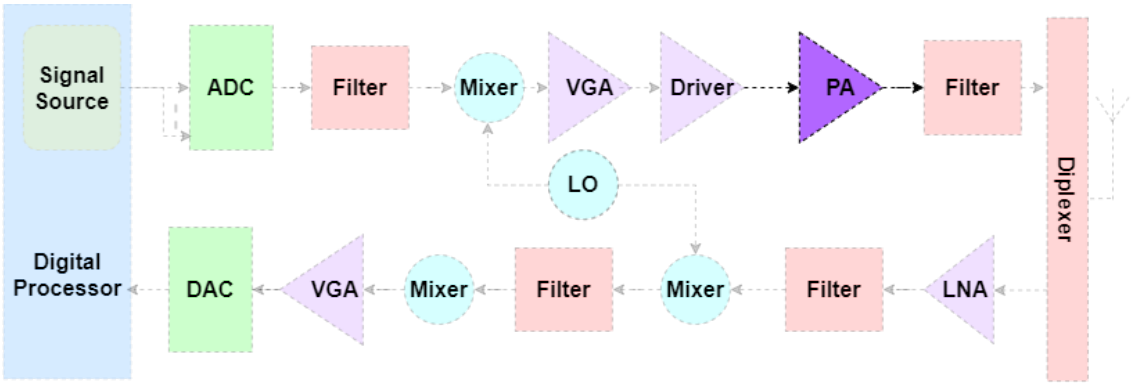

Figure 3.1: Components of concern for DPD

Figure 3.2 illustrates the nonlinear behaviour of a PA. A single carrier 5MHz WCDMA signal, sent through a PA at different power levels, the red signal is linear but trans-

mits at a much lower power, reducing its efficiency. The blue signal is transmitted at a much higher power level and thus experiences distortion. Nonlinear behaviour of a PA is typically more beneficial than linear behaviour as a better power efficiency is achieved, i.e. the output power of a PA can be maximised with minimal DC power. PAs that emit very low power behave inefficiently as a higher input power is needed to achieve a higher output power. In Figure 3.2 it can be seen that the AMAM curve is linear in the instance of a lower transmit power, in red, and nonlinear in the case of a higher transmit power, in blue.

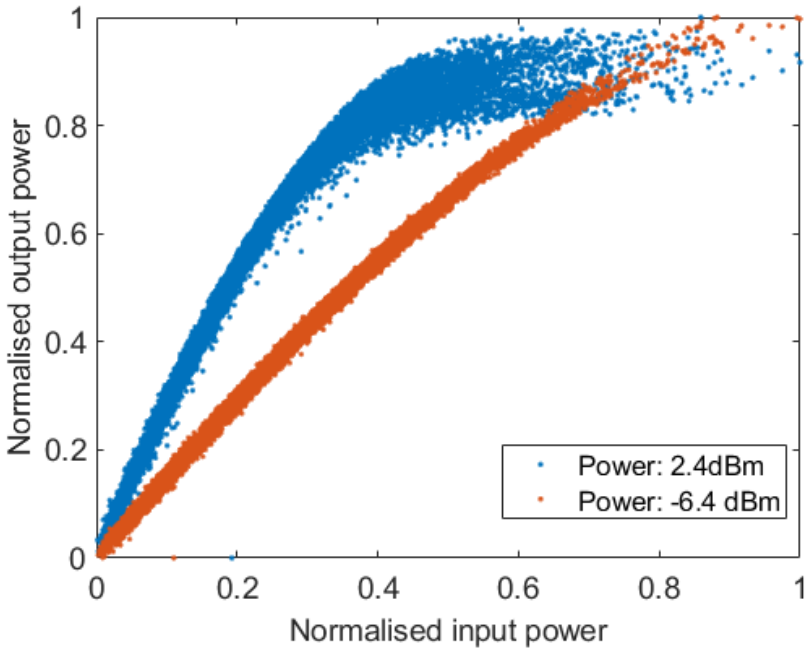

Figure 3.2: 5MHz WCDMA signal transmitted through a PA at different input powers to depict the nonlinearity elicited

Without compensation however the PA output will be nonlinear. The nonlinear behavioural generates spectral regrowth and could ultimately interfere with neighbouring channel signals. Spectral regrowth can be seen in plots of power spectral density, as the magnitude increase in the side lobes of the signal.

3.1 Power Amplifier Classes

In order to appreciate the division of RF PA's into classes, first the time domain behaviour of the active circuit must be examined. Average power is given in the time domain by equation (3.2), where V_p and I_p refer to the peaks of voltage and current respectively.

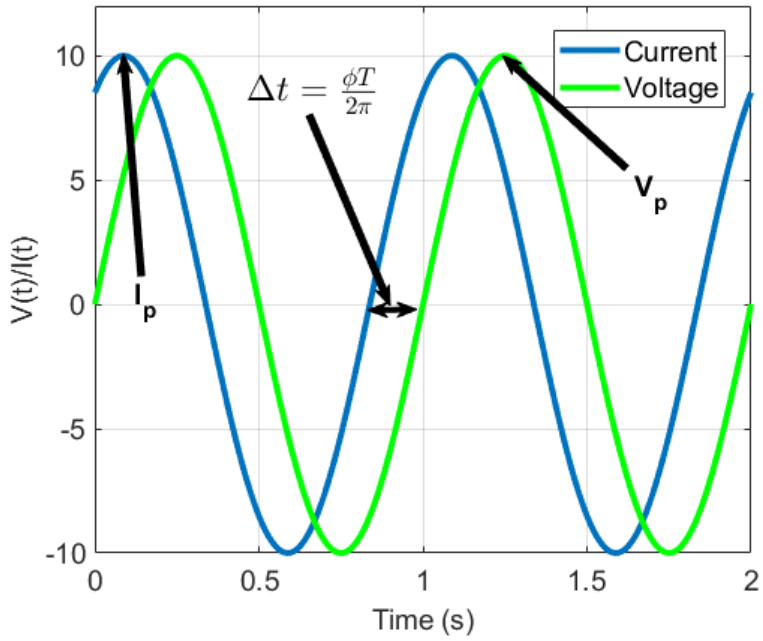

Figure 3.3: Voltage and current waveforms in time domain

$$P_{avg} = \frac{1}{T}\int_0^T v(t).i(t)dt \tag{3.1}$$

Where, $v(t) = V_p \sin(\omega(t))$ and $i(t) = I_p \sin(\omega(t) + \phi)$ and can be viewed in Figure 3.3.

The resultant of equation (3.1) can be viewed in Figure 3.4. Equation (3.1) can also be written discretely by equation (3.2).

$$P_{avg} = \frac{1}{2}V_p I_p cos\phi \tag{3.2}$$

The phase difference between the voltage and current waveforms, gives a minimum

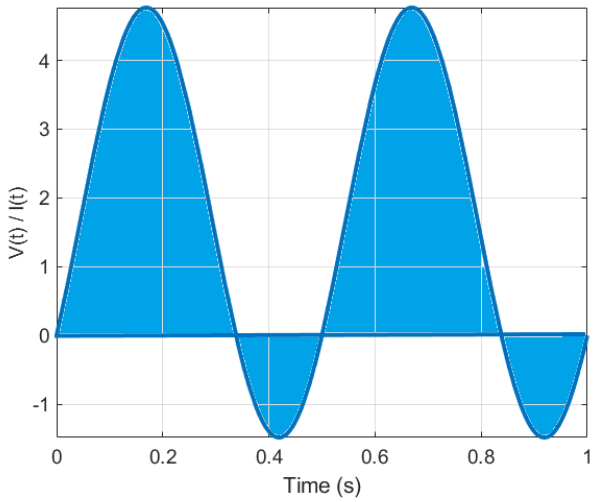

Figure 3.4: Equation (3.1), the integral of the product of $v(t)$ and $i(t)$

and maximum of power due the cosine function in the equation. At the maximum point of ϕ, the voltage and current are in phase, meaning power is being dissipated, converting energy into heat. At the minimum point of ϕ power is generated as the current and voltage waveforms are out of phase, and can be seen in Figures 3.5 and 3.6.

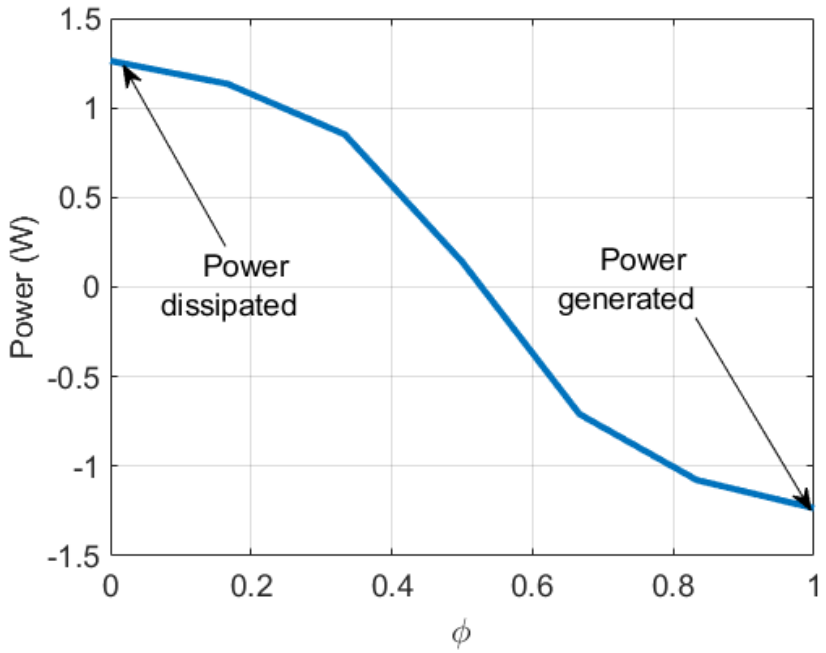

Figure 3.5: Power being generated and dissipated as $\phi = 0 to \pi$

In a practical sense one can determine if a component is a power generator or a power dissipater by plotting the sign of the voltage and current product - if it is

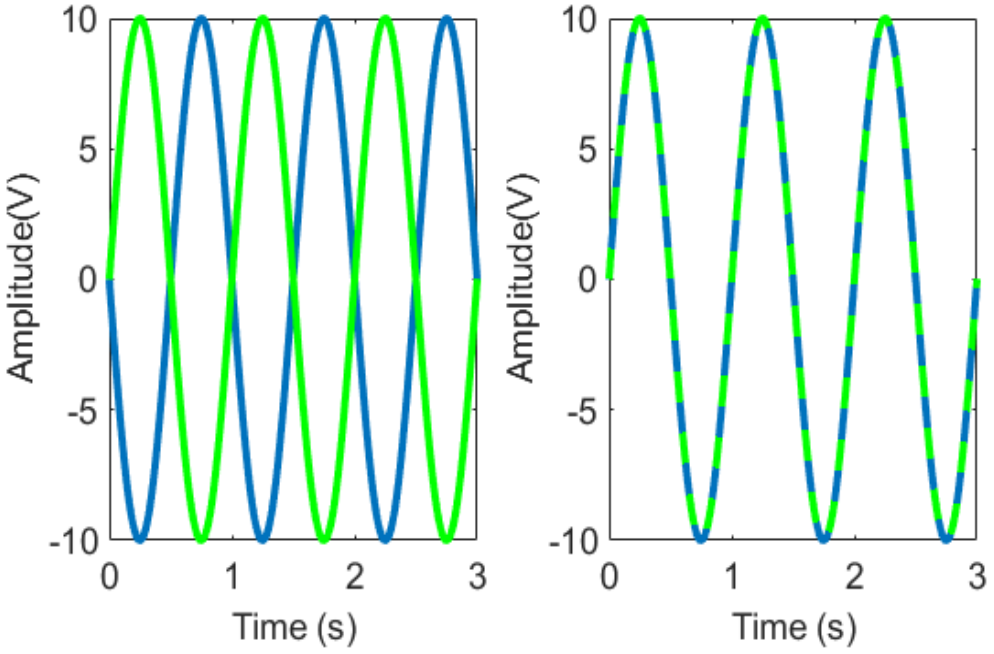

Figure 3.6: The Figure to the left shows that both the current and voltage waveforms are out of phase, i.e. power is generated. The Figure to the right shows that both the current and voltage waveforms are perfectly in phase i.e. power is dissipated

negative power is being generated. If it is positive power is being dissapated as seen in Figure 3.7.

PA's are classed based on their conduction angle, or ϕ, during which the PA passes current i.e. if the PA is constantly passing current, the conduction angle is 360°. The current and voltage waveforms are out of phase.

3.1.1 Class A Power Amplifier

Class A PAs allow current to flow for the entire cycle of the AC input supply i.e. the conduction angle is 360°. The complete signal present at the input is amplified at the output. The following Figure 3.8 shows an example of a circuit diagram for Class A PA, where the bias level depicts the DC offset originating from the transistor power source.

The function of a PA is to deliver RF power. Using a transistor such as a BJT or MOSFET powered by DC to produce a gain, amplifies the input signal by the gain to produce a higher power output signal. In the case of the class A PA the transistor

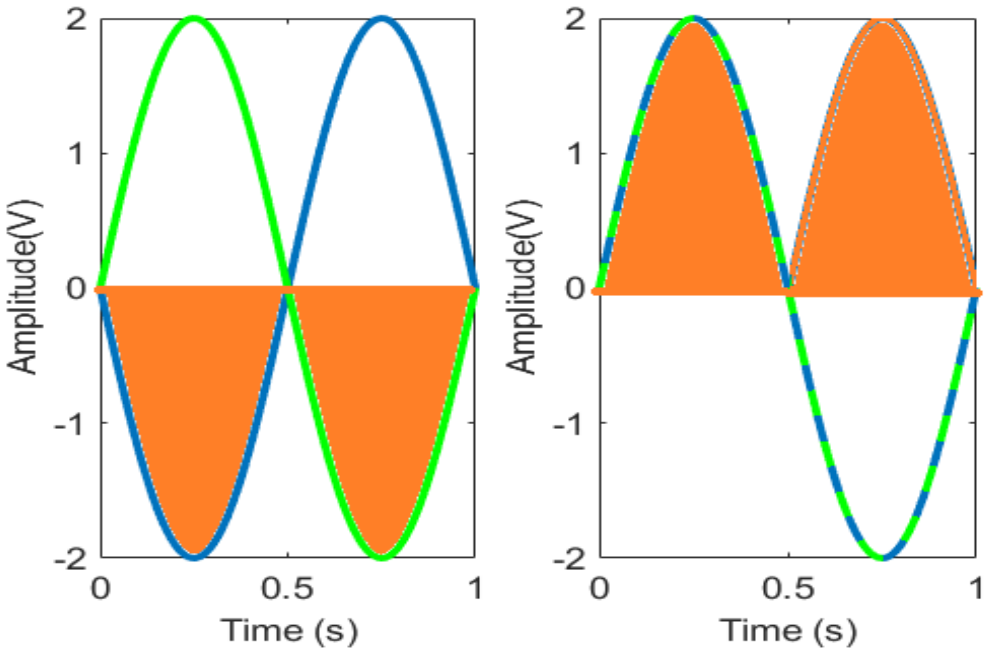

Figure 3.7: Left: Components generating power. Right: Dissipating power

is constantly biased in order to facilitate the conduction angle 360°, ensuring it conducts during one complete cycle of the input signal waveform. Class A amplifiers voltage and current waveforms are out of phase as in Figure 3.6.

Figure 3.9 shows a class A PA output waveform offset by a DC supply (i.e a bias level of DC current I_{dc} and voltageV_{dc}). In the case of Figure 3.9, the voltage and current wave forms, given by equations (3.3) and (3.4) respectively, fluctuate in value from approximately twice the V_{dc} and I_{dc}, to zero.

$$v(t) = V_p \sin(\omega(t)) + V_{dc} \tag{3.3}$$

$$i(t) = I_p \sin(\omega(t) + \phi) + I_{dc} \tag{3.4}$$

The dissipated power is highlighted purple in Figure 3.9, and given by equation (3.5), similar to that of equation (3.1).

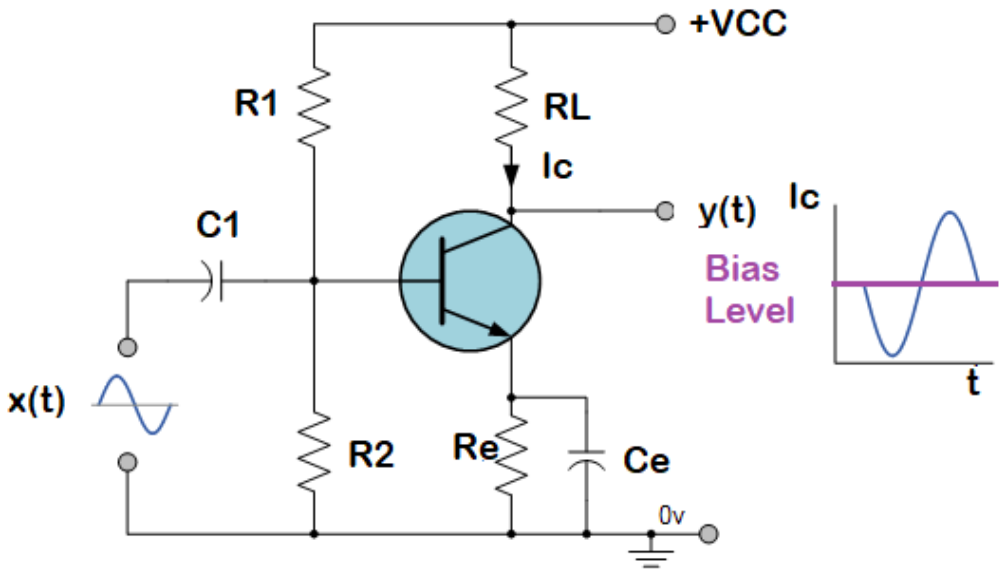

Figure 3.8: Example of class A PA circuit

$$P_{diss} = \frac{1}{T}\int_0^T v(t).i(t)dt \tag{3.5}$$

The power generated is calculated by removing the DC current and voltage biases from equation (3.5) and can be written as equation (3.6).

$$P_{gen} = \frac{1}{T}\int_0^T [v(t) - V_{dc}].[i(t - I_{dc})]dt \tag{3.6}$$

The result of equation (3.6) is negative and seen to be highlighted green in Figure 3.9. As energy must be conserved, the area of the dissipated power combined with the area of the generated power must be equal to the DC power supplied to the transistor, as given by equation (3.7).

$$P_{dc} = P_{diss} - P_{gen} \tag{3.7}$$

The output efficiency (η_{out}) of a PA is given by

$$\eta_{out} = \frac{P_{gen}}{P_{gen} + P_{dis}} = \frac{P_{gen}}{P_{dc}} \tag{3.8}$$

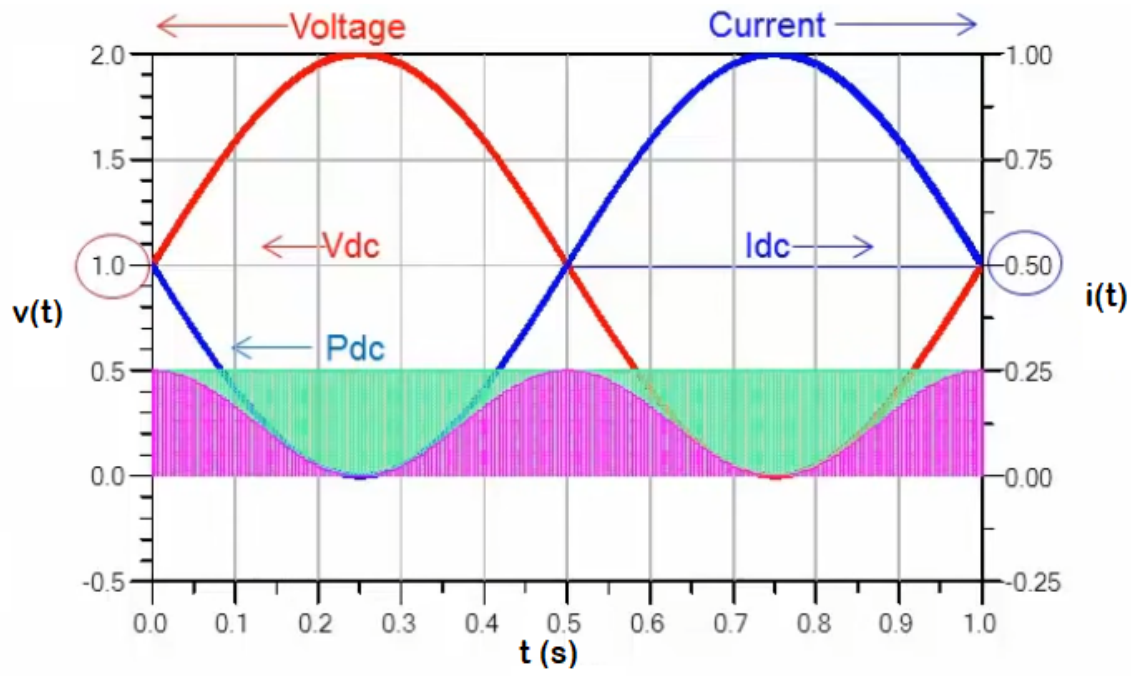

Figure 3.9: Class A PA waveform example, where V_{dc} and I_{dc} are circled

Table 3.1: PA classes and conduction angles

PA class	**Conduction Angle**
A	360°
AB	180° $\rightarrow$ 360°
B	180°
C,D,E,F	$<$ 180°

It can be deduced from equation (3.8) that for class A PAs half of DC power will be converted to RF power.

Classes of PAs are classed by their conduction angle, similar architectures are used as the above Class A. Table 3.1 displays commonly known PA classes and their corresponding conduction angle.

Classes of PA's have been well studied, as has efficiency enhancements using external circuitry by a number of researchers [13–17] and [18] .

3.1.2 Switch mode power amplifier: Class D

As articulated above, the drain efficiency of a PA can be significantly improved by reducing the conduction angle. In extreme cases, or switch mode PAs, the con-

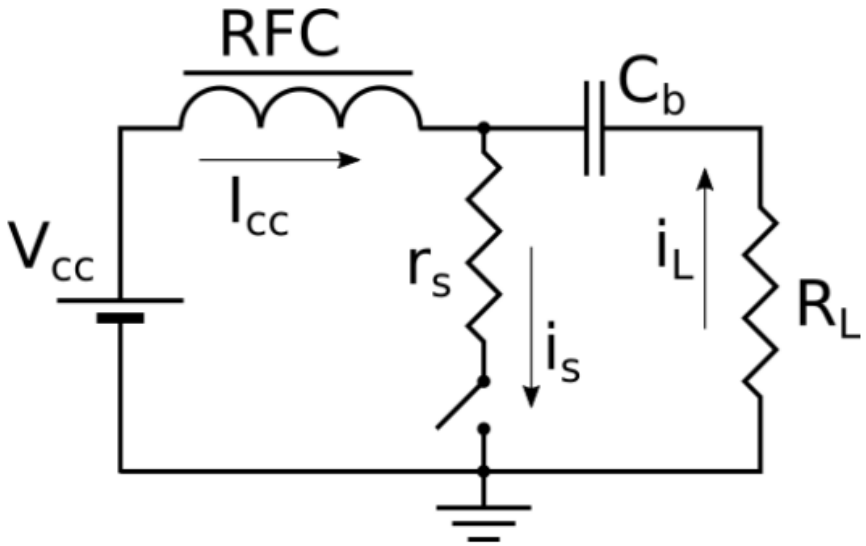

Figure 3.10: Ideal class D PA

duction angle is minimized and the PA device 'switches' between saturation and cutoff points. The saturation state refers to when the switch is turned on, i.e. the resistance is zero. The cutoff points refer to when the switch is off and experiences infinitely large resistance.

Class-D PA

The class-D PA was proposed in [26] to efficiently convert DC energy to AC energy. Since then this type of PA has been widely used for numerous applications, such as dc-dc converters, fluorescent lamps, and wireless transmitters, etc.

The prominent advantage of adopting class-D PAs, specifically for RF applications, is that of high drain efficiency compared to class-A, AB, and B PAs, due to the switch operation described above. There are two main categories of class D PAs:

- current switching : the drain is fed by a current source
- voltage switching : the drain is fed by a voltage source

The class D can be described with the aid of Figure 3.10. The transistor is represented as an ideal switch with a saturation resistance r_s. DC current is supplied from V_{cc}, RFC depicts an RF Choke. R_L is a part of the output circuitry as an

active load output connected to a DC block capacitor C_b.

Assuming the input signal to the PA is a square wave with a 50% duty cycle. The operation of the class D PA can now be described mathematically in two states, when there is infinitely large resistance and the second with no resistance.

When the switch is turned on, Kirchoff's law states that

$$i_s(\omega t) = i_L(\omega t) + I_{cc} \tag{3.9}$$

$$V_{cc} = i_L(\omega t)RL + i_s(\omega t)r_s \tag{3.10}$$

Solving equations (3.9) and (3.10) a mathematical expression for the switch current can be seen to be given in equation (3.11).

$$i_s(\omega t) = \frac{V_{cc} + I_{cc}R_L}{R_L + r} \tag{3.11}$$

Thus when the switch is on, the current cannot flow and the drain current can be expressed as equation 3.12.

$$i_s = \begin{cases} 0, & 0 \leq (\omega t) \leq \pi \\ \frac{V_{cc}+I_{cc}R_L}{R_L+r} & \pi \leq (\omega t) \leq 2\pi \end{cases} \tag{3.12}$$

As the input signal is square, there will be infinite harmonics in the drain current. The first two harmonics can be expressed using the Fourier expansion

$$I_0 = I_{cc} = \frac{1}{2\pi}\int_0^{\pi} i(\omega t)d(\omega t) = \frac{V_{cc} + I_{cc}R_L}{R_L + 2r_s}$$

$\mathrm{I}_1 = \frac{1}{\pi}\int_0^{\pi} i_s(\omega t)\sin\omega t d\omega t = \frac{4I_0}{\pi}$(3.13)

The power delivered to the load at the fundamental frequency will be :

$$P_1 = \frac{1}{2} I_1^2 R_L = \frac{8}{\pi^2 \frac{R_L V_{cc}^2}{R_L + 2r_s}^2} \tag{3.14}$$

The drain efficiency can now be determined as:

$$\eta = \frac{P_0}{I_O V_{cc}} = \frac{8R_L}{\pi^2 (R_L + 2r_s)} \tag{3.15}$$

To find the ideal, or most efficient case, $r_s = 0$. The drain efficiency of the class D PA would be $\eta = 81\%$, meaning that, 19% of the DC power supply is spent for higher harmonics. The main advantage of using the class-D PA is that there is a low voltage across the drain-source transition.

3.1.3 Doherty PA

The Doherty PA was first published in 1936 to ensure high efficiency for signals of PAPR between 6 to 10dB [19]. As seen in Figure 3.11 [20], the Doherty PA architecture is designed to implement a topology using two active devices. The active devices amplify components of the input signal conditional on the magnitude of the signal. Active load modulation is the overarching technique used by Doherty PAs which varies load impedance to PAs according to the transmitted signal's output envelope, allowing for a controlled variation of the instantaneous output power [21].

The Doherty PA is capable of amplifying signals with high PAPR, due to its two PA architecture, whilst maintaining high power efficiency. The main PA is typically a Class B or AB PA, sometimes referred to as the carrier amplifier.The auxiliary PA, sometimes referred to as the peaking amplifier, is usually a Class C amplifier [22]. Two quarter-wave transformers are connected at the input to the auxiliary PA and the output of the main PA, to compensate for a 90^0 phase shift introduced by the transformer in the main amplifier.

The two PAs are:

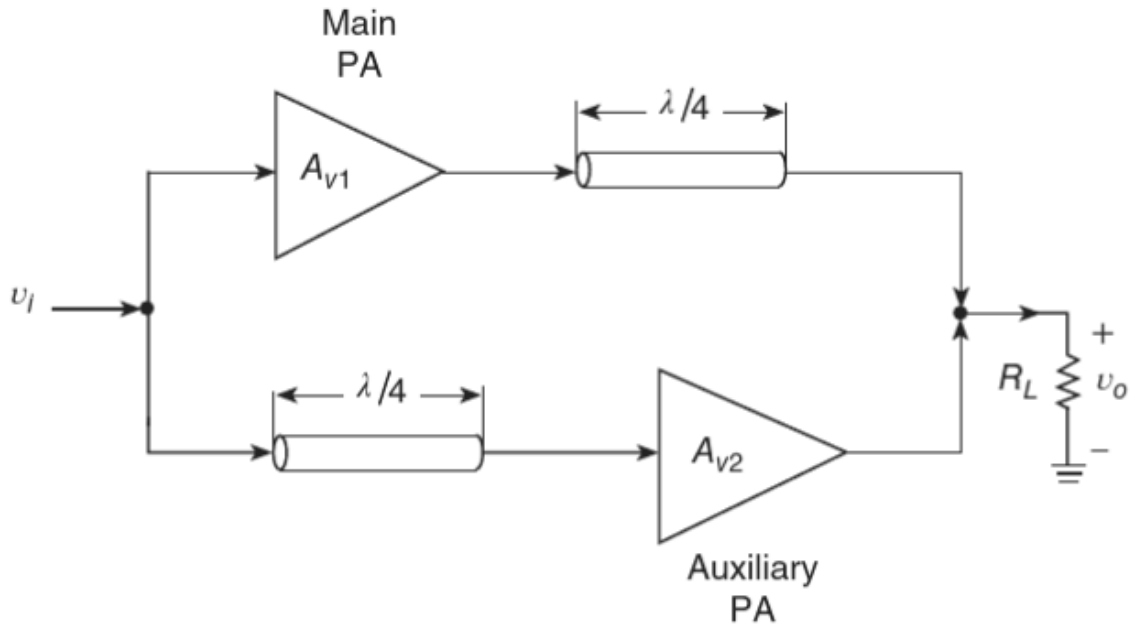

Figure 3.11: Block diagram of Doherty PA

- Carrier Amplifier : The carrier amplifier is typically operated in class B and induces gain at power levels at reduced powers compared to that of the peaking amplifier.

- Peaking Amplifier : The peaking amplifier is typically operated in class C. When the peaking amplifier is cut off the load impedance of the carrier amplifier is greater than that of the conventional amplifier, allowing the saturation region to be realised at a lower input voltage.

  When the peaking amplifier is conducting, it starts to generate an output current, increasing the impedance seen from the carrier PA. In order to reduce the power generated by the carrier PA to the load, an impedance inverter is added. The impedance variation in the time domain dictates that the carrier amplifier remains in the more efficient saturation mode as the power generated by both the carrier and peaking PA's increase . If the current of the peaking PA increases twice as fast as the current of the carrier PA, at the peak value of the input signal they both generate the same peak power [23].

As in the class A PA above, the input voltage of the source dictates its output current due to Ohm's law.

The Doherty PA, as seen in Figure 3.11, requires configuration of a splitter and combiner. The splitter and combiner allow the power to be directed to the corresponding amplifier and the outputs to be added to provide the complete output. Both the splitter and combiner need to accomplish phase and matching requirements to ensure maximum power efficiency [24].

There are different configurations of the Doherty amplifier

- Symmetric Doherty: the peaking and carrier amplifiers have the same size. Both amplifiers contain identical matching circuitry and input voltages. The symmetric Doherty suffers from power degradation because the bias point of the peaking PA is lower than that of the carrier PA. The current of the peaking PA at maximum input drive voltage reaches below the maximum current level, i.e full modulation of the load is never achieved.

- Asymmetric Doherty: In order to overcome the power performance of the symmetric Doherty PA, Asymmetric Doherty PA architecture is altered by applying more power to the peaking cell. The current and voltage maximums, as described above for the Class A PA, of the peaking amplifier increase proportionally to the input voltage level. The voltage reaches the maximum voltage at the maximum input voltage [24]. The main motivation for using an Asymmetric Doherty PA is to increase the peak to average power backoff.

- Digital Doherty : the digital Doherty PA configuration allows the input signal to be split digitally and fed to both the peaking and carrier PAs separately. The digital Doherty PA enables performance improvement by allowing for adaptive phase alignment in the digital domain, which compensates the phase impairment between the two paths when the peaking amplifier is active. The digital Doherty eliminates phase delay inherent to impedance mis-matching.

- N way Doherty : The N way Doherty PA configuration provides N parallel paths with N-1 peaking amplifiers. The topology of the N way Doherty PA allows for higher efficiency i.e. achieving higher output power at back off

regions. The N way Doherty is an active area of research [19, 24] with new topologies released for specific use cases.

3.2 Power Amplifier Behavioural Modeling

Behavioural modeling is used, in this thesis, to produce models of PAs that take into account nonlinear distortions and memory effects and characterise them adequately while maintaining high accuracy. Behavioural modelling is a mathematical approximation of the system's response to a stimulus. A PA's behaviour can be estimated by examining system input/output relationships. PA behavioural models are commonly constructed using polynomial equations. Polynomial models used to characterise PA behaviour are often capable of being used as DPD structures for linearisation.

Behavioural modeling is an active area of research as it is difficult to enscapulate a PA behaviour, one model can be derived for one operating point, but should the input magnitude change the model may no longer be accurate. To compound this problem heat may also effect PA behaviour [25], other sources of distortion can also be present such as coupling [26].

3.2.1 Memoryless models

Memoryless behavioural models do not take into consideration previous values of input/output signals, assuming the current output relies on the current input. Memoryless models can only model nonlinearities that have little to no memory effect.

The static traits of a PA amplitude to amplitude conversion (AM/AM), as seen in Figure 3.2, and semi static traits of amplitude to phase conversion (AM/PM) are assumed to be static, allowing the model to consider the output signal as a function of the phase and amplitude of the input signal.

The memoryless polynomial model is a simple yet powerful technique commonly

used for PA behavioural modeling [27]. The model characterises a range of input amplitudes, measuring the output amplitudes and phases, such that the response of the system can be determined as in original works by [28].

3.2.2 Look Up Tables

Look-up-table (LUT) solutions for nonlinear power amplifiers have been previously introduced for both behavioural modelling and predistortion. The chronology of these LUT solutions shows new methods have emerged for both applications.

As a means of modeling the behavior of RF power amplifiers, a number of different LUT approaches have been proven to be effective. A data-based nested LUT structure has equivalent performance to the memory polynomial for modeling of power amplifiers exhibiting memory effects as shown in [29].

This LUT approach has been further extended to a 2-D LUT model for transmitters/PAs exhibiting memory effects. With an additional dimension [30], the LUT is expanded to take into account the dependency of the device behavior on the preceding samples. The 2-D LUT models the transfer function of the device under test as a complex gain that is a function of the magnitude of the current and previous samples.

More recently, Nunes et al. have demonstrated a LUT solution for high efficiency power amplifier architecture [31].

3.2.3 Nonlinear behavioural models with memory

Nonlinear models with memory terms can account for nonlinear memory effects as required. The inclusion of nonlinear memory computations allows for a more accurate behavioural model, as effects from nonlinear memory are accounted for. Nonlinear models with memory are focused on in this work.

Volterra Model

A model structure that encompasses memory effects and non-linearities is the Volterra series. Comparing the Volterra model to other more compact models, such as memory polynomial and generalised memory polynomial, it was found that the Volterra model gives the most accurate modeling performance [32].

Behavioural modeling using the Volterra series combines numerous linear convolution and a nonlinear power series, allowing the system to be modelled while incorporating memory effects [33]. Although the accuracy of the Volterra model is high, the computational complexity is also high as the number of parameters to be estimated escalates rapidly as the nonlinear order of the model or the memory depth are increased.

A system with finite order of non-linearity with finite memory depth can be described in the discrete time domain, for a real value signal, by equation (4.1).

$$y(n) = \sum_{p=1}^{P} y_p(n) \tag{3.16}$$

Where,

$$y_p(n) = \sum_{i_1=0}^{N-1} \cdots \sum_{i_P=0}^{N-1} h_p(i_1, \cdots, i_p) \prod_{i=1}^{p} x(n - i_r)$$

Where $x(n)$ and $y(n)$ are the input and output signal to the system respectively. $h_p(i_1, ..., ip)$ represents the filter co-efficient expansion using, p, the highest order for the non-linearity of the Volterra series expansion. N represents the maximum memory tap length [34].

Please note only the odd powered terms are retained for the Volterra model and even terms are discarded since they do not appear in the passband response if the input signal is complex [35].

Pruned Volterra Models

Subset model structures are often used in practice to reduce complexity while still preserving high linearisation capabilities.

To perform modelling of a nonlinear channel using complex valued quadrature signals, the complex baseband Volterra series was first introduced and derived in [36]. Benedetto et al. proved in this work that odd-order passband Volterra series terms produced outputs centred at the carrier frequency.

Inspired by [36], research was initiated by pruning even-order terms from many polynomial baseband models, such as works [37–41]. By excluding the even order model coefficients authors were able to reduce model complexity with minimal loss of accuracy/performance .

The linearisation performance improvement gained from the inclusion of even-order terms has been analysed in the works [42, 43]. The inclusion of even order terms can provide improved linearisation capability, [42, 43], thus exclusion of odd terms should be considered on a case by case basis.

Dynamic deviation reduction, [44], proposed reducing the number of memory terms at each order of nonlinearity of the model. The authors established that the effects of the nonlinearities decrease as the nonlinear order increases.

Radial pruning of Volterra series, proposed by [45], allows for pruning to be performed based on a terms radial direction. Radially pruned Volterra is expressed Volterra kernel mathematically as cubes and computed as (3.17).

$$y_{\mathrm{RPV}}(k) = \sum_{n=1}^{N} \sum_{r=1}^{n} \sum_{s=1}^{S_{n,r}} \sum_{q=0}^{Q} h_{n,r,s}(q) \Pi_{n,r,s} \left\{ x(k), q \right\}. \tag{3.17}$$

Where r is the radial direction, s is the number of radial directions for a given Volterra kernel and q is a scalar delay.

Polar and orthogonal descriptions of the Volterra series have been investigated [46].

Pruning strategies are currently an active area of research. Producing concise models allows for many potential benefits such as - lower latency and lower computational cost.

Memory Polynomial Model

A polynomial constructed from the product of Volterra terms and terms characterising the signal envelope is known as a Memory Polynomial (MP) [42].

The MP is a model derived from the Volterra model comprised only by products with the same time-shifts [47], and can be thought of as the diagonal coefficients of the Volterra Model coefficient matrix. This model is formed by combining selected individual filters for each nonlinear order. Combining these individual filter responses is given by equation (3.18)

$$y(n) = \sum_{p=1}^{P} \sum_{m=0}^{M} a_{pm} x(n-m)|x(n-m)|^{p-1} \tag{3.18}$$

where a_{pm} are the estimated model parameters. P and M represent the highest nonlinear order and the memory depth of the model, respectively.

The MP model employs P $\cdot$ M coefficients [48]. Subset models of the MP do exist, such as the Envelope Memory Polynomial (EMP) [49].

The memory polynomial presented in equation (3.18) is an efficient alternative method of modelling non-linear PAs. Comparing MP to Volterra series to model nonlinear PAs the MP model is computationally less complex but can be less efficient [32].

Generalised Memory Polynomial Model

The Generalised Memory Polynomial (GMP) [50] can be considered as MP models, given by (3.18), to include leading and lagging cross-terms as seen in equation (3.19).

$$
\begin{aligned}
u = & \sum_{m=0}^{M} \sum_{k=1}^{K} h_{mk} \cdot x(n-m) \cdot |x(n-m)|^{k-1} \\
& + \sum_{m=0}^{M} \sum_{k=2}^{K} \sum_{p=1}^{P} h_{mkp} \cdot x(n-m) \cdot |x(n-m-p)|^{k-1} \\
& + \sum_{m=0}^{M} \sum_{k=2}^{K} \sum_{q=1}^{Q} h_{mkq} \cdot x(n-m) \cdot |x(n-m+q)|^{k-1} .
\end{aligned} \tag{3.19}
$$

K_a and L_a index the arrays for the input signal and it's envelope;K_b, L_b and M_b refer to the indexing of the input signal and it's lagging envelope; and K_c, L_c and M_c and are index arrays for the input signal and it's leading envelope.

a_{kl}, b_{klm} and c_{klm} are the estimated model parameters. The cross-term basis functions consist of signal permutations of input signal samples multiplied by the complex envelope of delayed samples, as seen in equation (3.19). Subset models of the GMP have been proposed such as the Complexity Reduced Generalised Memory Polynomial (CR-GMP) [48].

Both the MP and GMP consist of a reduced number of coefficients compared to the Volterra series. The Volterra Series is more accurate in describing nonlinearities [32] when compared to GMP and MP but is also requires much more computation.

3.3 Power Amplifier Linearisation techniques

Numerous techniques to linearise the distortion of PAs have been proposed. These techniques can be roughly divided into three categories.

- Feedback
- Feedforward
- Predistortion

Both feedback and feedforward techniques were introduced by H.S Black in the early

1920's [51].

Feedback linearisation techniques operate by utilising the output signal of the PA and subtracting it directly from the proceeding input signal to the PA, ensuring linear behaviour. This can be seen in block diagram form in Figure 3.12 .

3.3.1 Feedback technique

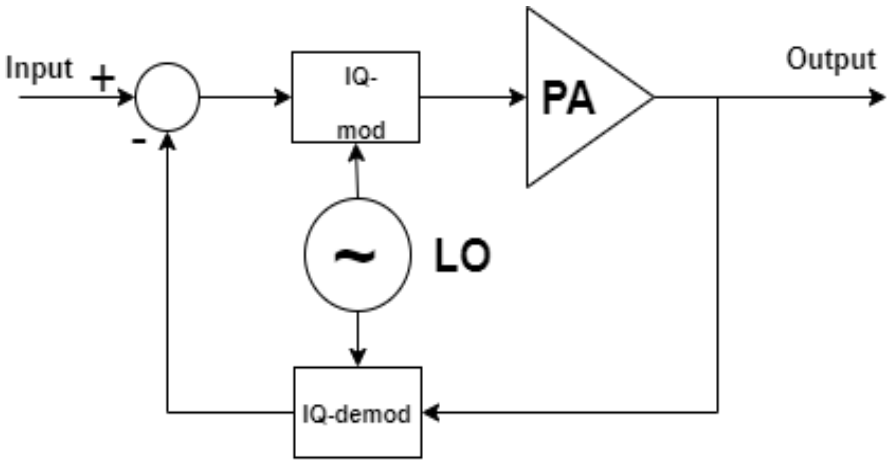

Figure 3.12: Feedback linearisation technique

Feedback linearisation techniques can reduce nonlinearity but suffers from potentially high latency due to the architectural delay, limiting its application to narrow band signals. The update is applied at the input to the PA. Feedback linearisation has been widely employed in the past to provide compensation as it is typically more efficient than feedforward techniques. Feedback linearisation can be achieved by various mathematical procedures including polar, Cartesian and RF [15].

3.3.2 Feedforward technique

Feedforward linearisation uses two amplifiers. Distortion from the nonlinear PA output (PA_m) is extracted and subsequently transmitted through a second error amplifier (PA_e). Transmitting through the error amplifier allows the error to be amplified and subtracted from the signal PA_m resulting in a linear output signal (y).

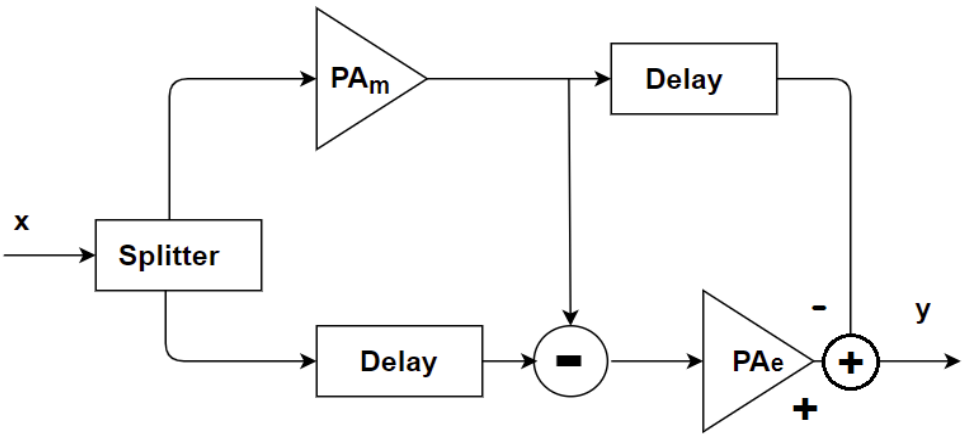

Figure 3.13: Feedforward linearisation technique

Linearisation performance utilising feedforward techniques can be advantageous but the net power efficiency achieved is typically low due to power consumption costs of the error amplifier [52, 53].

3.3.3 Predistortion technique

The pre-distorter technique manipulates a PA input to mitigate against nonlinearity introduced by the PA. The PAs gain is nonlinear, thus the pre-distorter correction is also nonlinear. Predistortion techniques have improved greatly over recent years as digitally applied predistortion has become more advanced due to increasing computational power. Digital predistortion (DPD) can be seen in Figure 3.14. Analog predistorton is also achievable but less commonly implemented. Analogue pre-distorters use active devices such as field effect transistors and/or passive devices, for example diodes, to perform linearisation but analogue devices are inherently less flexible than digital implementations [15] and more costly.

DPD is implemented as a predisorter block before the PA. The digital domain is used to introduce a complimentary or opposite nonlinearity caused by the PA. The complimentary or inverse function of the nonlinearity is determined using various DPD techniques, such as adaptive basis function polynomials, and will be the main area of research for this work.

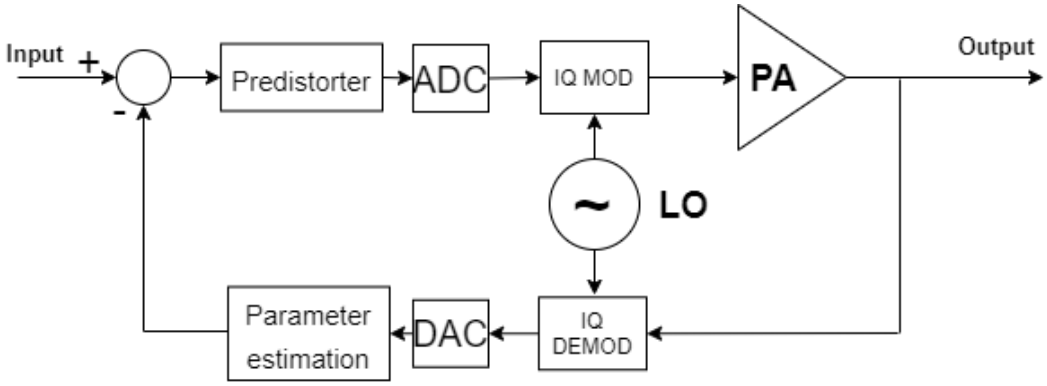

Figure 3.14: Digital predistortion linearisation technique

DPD can provide a desirable solution to illicit linear behaviour and is a good compromise in terms of performance and computational complexity. Behavioral modeling of PAs requires functions that describe dynamic nonlinear behaviors, while DPD synthesis is establishing functions of the inverse of the behaviour.

DPD, as depicted in Figure 3.14, presents the challenge of determining a predistorter function, such that the output of the predistorter and PA system achieves highly linear relationship to the input. The challenge lies in selecting a model structure to implement the predistorter function and to determine the parameters of such a model.

For the purpose of this work we are concerned with two DPD parameter identification techniques.

- Indirect learning architecture (ILA)
- Direct learning architecture (DLA)

Indirect learning architecture

ILA is a parameter identification technique that was introduced by [54] and adapted for use in DPD by [55]. ILA utilises the inverse modeling approach. A post-inverse model, the postdistorter, of a PA is identified by using the PA output signal $y(n)$ to model the PA input $u(n)$ and depicted by Figure 3.15.

Once the postdistorter has been identified, the describing parameters are then repli-

cated and used as the predistorter.

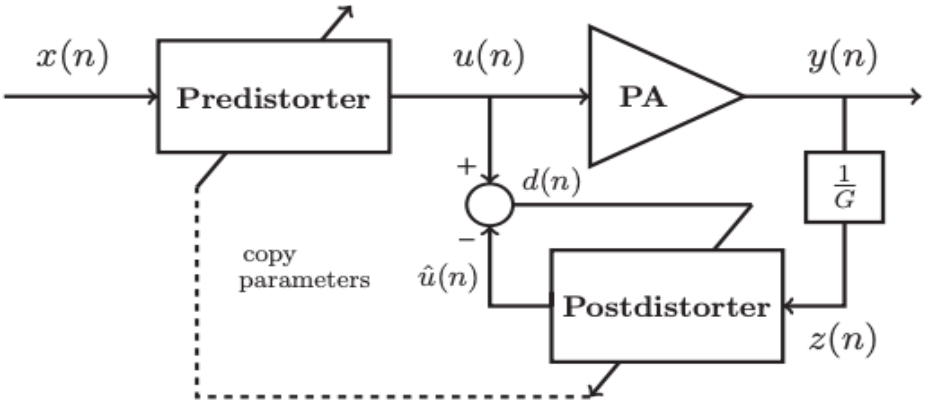

Figure 3.15: Indirect learning architecture

The ILA criteria for minimisation can be expressed:

$$\hat{\Theta} = \underset{\theta}{\operatorname{argmin}} \|d(n)\| \tag{3.20}$$

where $d(n) = u(n) - \hat{u}(n)$, the error between the predistortion function and post-distorter function, respectively. ILA assumes the error $d(n)$ converges to zero [56], leading to the assumption:

$$y(n) = Gx(n) \tag{3.21}$$

Where G represents gain or scaling factor to normalise the power output of the PA $y(n)$ to the power of the input signal to the predistorter input signal $x(n)$.

A problem with ILA is the error $d(n)$ does not converge to zero. There are many reasons for this, for example:

- The predistorter model may not be appropriately structured
- The parameters of the predistorter model may not be estimated without introducing error
- Incorrect memory depth or nonlinear order of the model may not be suited to

the input/output signal characteristics.

A study carried out in 2012, [57], showed that utilising a least squares error based ILA parameter identification technique the presence of measurement noise occurring in $y(n)$ causes the parameter estimates to converge towards a biased solution.

Another problem to be considered for ILA is that of gain normalisation, determining G, from equation (3.21). There are many proposed solutions to calculating G [58–60], all of which cause different effects to the output power of the PA. Although [58–60] all achieved high linearity, the altered average power of the PA before and after DPD is undesirable.

Direct learning architecture

DLA formulates a pre-inverse model of a PA, directly diminishing the error between the output of the PA, $z(n)$, and the estimated output signal, $\hat{z}(n)$ such that:

$$e(n) = \hat{z}(n) - z(n) \tag{3.22}$$

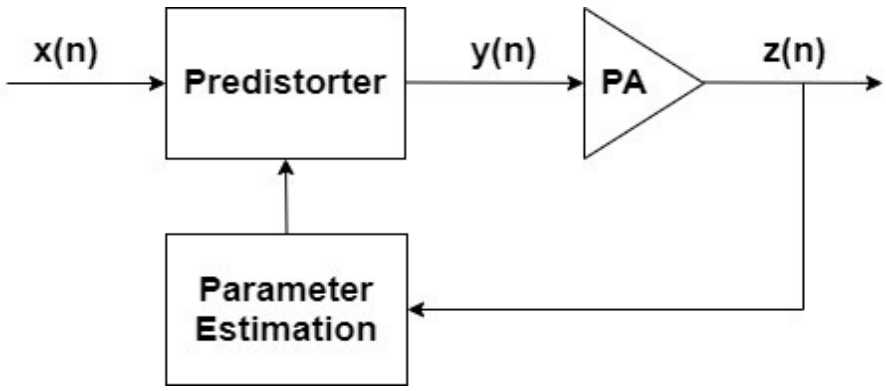

Figure 3.16: Direct Learning Architecture

DLA estimates the parameters of the predistorter utilising digitally implemented algorithms, depicted by Figure 3.16 as the Parameter Estimation block. There are many proposed methods for DPD using DLA [61, 62]. Historically, an argument against the use of DLA was that the algorithms were too computationally cumbersome, but as technological advances continue, modern digital signal processors and

FPGAs are increasingly improving relative to the latency experienced in the past.

DLA is implemented by identifying a first iteration forward model of the PA, an algorithm, which models the nonlinearity is used to estimate the predistorter parameters that minimize the error, seen in equation (3.22), which are then in turn used to generate a predistorted signal $u(n)$ that is applied to the real PA at its input.

Comparative studies have been carried out to determine the best parameter identification technique, ILA or DLA, to utilise [63, 64] but as stated, these comparisons may be out of date as in recent times [65] adaptive DLA and ILA with high computational complexity have been implemented with low latency. The work in this thesis uses both DLA and ILA in different instances and a comparison of the methodologies proposed using both ILA and DLA is being considered.

3.4 Power Amplifier Modeling and Linearisation Routines

3.4.1 Least Squares

Least squares (LS) estimation is a process where the input and output signals are analysed in vector form. The optimal solution is derived from direct matrix inversion [46]. The LS solution is identical to the solution of the minimum mean square error (MMSE) solution, although is dissimilar in the method of extraction. The LS solution can be determined solely using the input and output data. The LS algorithm determines coefficients such that the cost function given by (3.23) is minimised.

$$J(n) = \frac{1}{n}\sum_{t=1}^{M}(y(t) - X^T w)^2 \tag{3.23}$$

Where y is the output signal, X the input signal and w the coefficients.

The most advantageous weights are guaranteed to be determined if the data set is representative of the device under test. The disadvantage of the LS estimation

process is that if X is large the computational complexity is inherently larger also.

3.4.2 Recursive least squares

Recursive least squares (RLS) is a well known training method used for training the behavioural models of PAs [66, 67]. The exponentially weighted RLS algorithm can be adequately described in terms of its cost function. Model coefficients, are adapted based on the cost function $J(n)$, shown below in equation (3.24).

$$J(n) = \sum_{k=1}^{n} \lambda^{(n-k)}(d(k) - H^T(n)\boldsymbol{X}(k))^2 \tag{3.24}$$

λ is an exponentially weighted factor, $0 < \lambda < 1$, controlling the convergence speed of the function, referred to here, as the forgetting factor. λ closer to 1 enables the algorithm to decay slowly, tracking signal alterations more closely. The inverse is true for λ tending to 0. $d(k)$, refers to the actual output signal at sample k. Filter coefficients, $H(n)$, are determined such that the weighted average of the squared estimation error is minimised from time $k = 1$ to $k = n$. [46]. $\boldsymbol{X}(k)$ represents the input signal to the model at sample k.

Both the Recursive Least Squares (RLS) and Least Mean Square (LMS) technique have been employed in the calculation of DPD parameters [46, 68]. Detriments to these methods however, are slow convergence rates (LMS), due to over determined input/output relationships and resulting in intense computations [69] and instability without modification. Using RLS without limiting the input training signal length could also lead to instability during training [46].

3.4.3 Batch training - least squares

LS can be altered to be a batch estimation technique. Authors of [70] estimate that it is the Best Linear Unbiased Estimator (BLUE) once Gauss-Markov criterion are met. Best, according to the authors, meaning it provides the estimator with

minimal variance. Basis functions are derived from a block of samples and then used to extract DPD/behavioural modelling parameters. DPD/behavioural modelling coefficients, h , can be calculated using equation (3.25) [71].

$$h = (Y^T Y^{-1}) Y^T e \tag{3.25}$$

Y is a matrix of basis functions. Each matrix of basis functions describes a signal permutation of the PA output signal y and the error signal e.

3.5 Performance metrics

3.5.1 Normalised mean square error

A metric utilised to assess a models performance is Normalized Mean Square Error (NMSE). NMSE can be written as equation (3.26).

$$NMSE = \frac{\sum_{n=0}^{N-1} |y(n) - \hat{y}(n)|^2}{\sum_{n=0}^{N-1} |y(n)|^2} \tag{3.26}$$

$y(n)$ is the measured signal at the PA output and $\hat{y}(n)$ denotes the estimated modeled output. NMSE is used to evaluate the in-band performance of the model.

3.5.2 Error Vector Magnitude

Error Vector Magnitude (EVM) is another figure of merit utilised when measuring the accuracy of a model and is determined by the integrity of a digital modulation scheme utilised. Analysing the discrepancy between the ideal constellation points of a modulation scheme and the experimentally measured PA output signal is utilised to determine an error known as EVM. The EVM is defined by [72] given by Equation 3.27.

$$EVM = \frac{\sum |Y(k) - Y_d(k)|^2}{\sum |Y_d(k)|^2} \tag{3.27}$$

where $Y_d(k)$ is the constellation points extracted from the reference signal $y_d(n)$ after demodulation and $Y(k)$ is the constellation extracted from the measured output signal $y(n)$. EVM is used to estimate the impact of signal processing techniques, the non desirable effects in the transmission of the signal [73].

Error vector magnitude (EVM) is a measure of distortion caused by a nonlinear system. EVM is calculated by evaluating individual signal samples and comparing them to ideal constellation points. Ideal constellation points and measured constellation points can be seen in Figure 3.17, resulting of an average EVM of -10.2dB.

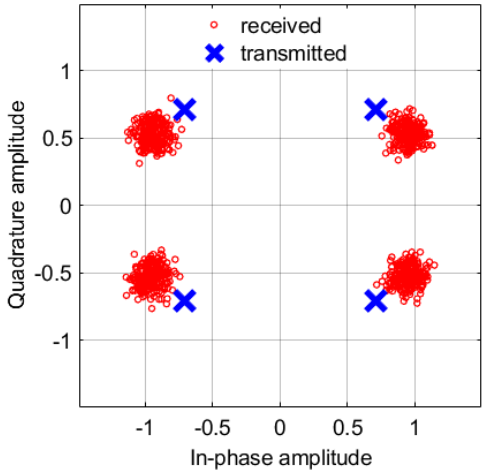

Figure 3.17: QPSK constellation diagram with visually appreciable EVM error. The input signal was transmitted through a channel injected with 20dBW Additive White Gaussian Noise (AWGN).

3.5.3 Adjacent channel power ratio

Adjacent channel power ratio (ACPR) is an out-of-band performance metric, measuring the power of the distortion components that leak into an adjacent channel compared to the power of the signal in the main channel [74]. The ACPR is given as equation (3.28).

$$\text{ACPR[dB]} = 10\log_{10}\frac{\text{P}_{\text{adjacentchannel}}}{\text{P}_{\text{mainchannel}}} \tag{3.28}$$

The adjacent channels are the frequency bands equal to the baseband signal bandwidth, either side of the main channel, in the frequency domain. Adjacent bands typically contain the Inter Modulation Distortion (IMD) products from a nonlinear PA.

3.6 Chapter summary

In conclusion, there is substantial work done and being done in the modeling and linearisation of multiple classes of PAs. The experimental works proposed in this thesis utilises both Doherty and Class AB PAs. Many researchers have proposed methods to simplify PA models, and metrics to quantify the savings i.e low NMSE indicates an accurate model. This thesis focuses on models deduced from the Volterrra series and Doherty class PAs. In further chapters both RLS and LS batch estimation are used as training techniques and both DLA and ILA are used in validating PA DPD technqiues.

Chapter 4

Ensuring Stability for DPD Training

A considerable drawback with using the Recursive Least Squares (RLS) technique is the instability of the coefficients during the training of the model. Using RLS without limiting the input training signal length can lead to instability during training [46]. This chapter provides a computationally efficient technique to detect the onset of instability during adaptive RLS training and subsequently to inform the decision to cease training of dynamic memory polynomial based behavioural models, to avoid incorrect training after the onset of instability. The proposed technique does not require modification of the RLS algorithm, merely an observation of the pre-exsisting autocorrelation function based update. This technique is experimentally validated using four different signal modulation schemes, LTE OFDM, 5G-NR, DVBS2X and WCDMA.

4.1 Related works

In this section of this thesis, a review of the existing works contributed towards RLS instability detection is presented with associated advantages and disadvantages.

Previous works by the author of [75], demonstrated that altering the RLS algorithm

to include periodic regularisation and maximum and minimum eigenvalue limitations can help to maintain stability. Regularisation by padding the diagonal of the autocorelation function causes a slight degradation in dynamic range. The method of padding the autocorrelation function is relatively simple although it is not disclosed in the paper how the eigen decomposition was calculated. The disadvantage of this methodology that extra computations are needed. Computations are needed to set the eigenvalue thresholds and to determine a suitable padding window. A disadvantage associated with this methodology is that the padding window is susceptible to noise and that disparate accuracy was reported dependent on the eigenvalue limits and regularisation period.

Authors of [76] alter the RLS algorithm by adapting a hybrid approach of directional and exponential forgetting factorisation to implement an adaptive forgetting factor. The proposed methodology ensures stability and convergence to a minimum error. Disparate accuracy of the proposed method is reported when comparing results using alternate PAs. The results are highly dependent on *a priori* statistical PA data. The high data dependency leads to reduced accuracy of RLS estimates. Due to the high computational complexity of works by [76], this method suffers from increased latency.

In [77] two computations of DPD coefficients are performed, with one set of coefficients specifically containing peaks. The computational complexity introduced by the necessity of computing two sets of DPD coefficients is an unattractive solution. Authors of [77] report an improvement in error estimation when compared to the RLS algorithm alone but it does not eliminate instability.

Research conducted by [78] proposes utilising a hybrid approach adopting both RLS and Least Mean Squares (LMS) for performing DPD. The adaptive algorithm utilises RLS when the error signal is large, for quick convergence, and subsequently automate their algorithm to adopt LMS when the error is below a set threshold value. The work by [78] details an improvement of 17dB when compared to a system without

DPD. The validation in this work is simulated. This hybrid approach for DPD produces favourable results, but a sudden increase/decrease in error, as experienced in typical experimental work, could cause the error threshold to switch RLS to LMS or visa versa.

Recent research [79] presents a method to modify the RLS algorithm by applying an error threshold. The update is based on the computation of the error as seen in (4.7). Should the value of the error at an instantaneous time sample be above the error threshold at any particular time sample, the algorithm re-computes the error for the following time sample, omitting the previous from the calculation of the DPD coefficients. The methodology proposed by [79] reports a 30dB improvement when compared to the un-modified RLS algorithm. The validation performed in [79], using a memory depth of 2, is simulated but the methodology would be robust against spurious noise. Authors of [79] present results which indicate severe latency, which indicates that this method would be unsuitable for use in practice by a base station, as very low latency is required by the 5G-NR standard [2] .

In [80], Swaminathan et al, propose a variable forgetting factor. Computing both a variable convergence factor and variable forgetting factor improves the steady state alignment of the proposed method when compared to both the LMS and RLS algorithm. [80] reported a 6dB mean square error improvement when compared to the RLS algorithm. Authors of [80] minimise the a priori error signal of the RLS function when determining the variable convergence factor and variable forgetting factor, which are updated by thresholding the value of the bit error rate. The authors present a look up table method to apply DPD.

Works by [75–79] and [80] all propose techniques that require additional computations and or a modification of the RLS algorithm in order to maintain stability during training. In this thesis a method is presented to produce a computationally efficient approach to avoid the onset of instability during model coefficient training for the RLS algorithm. The aforementioned works present alternative methods

Table 4.1: Training signal length relationship to memory tap length

Memory tap length	Maximum signal length (samples)
1	13400
2	11400
3	7800
4	7100
5	7000
6	5800
7	5200

Table 4.2: Training signal length relationship to sampling frequency

Sample frequency(MHz)	Max signal length(samples)	RMSE
7.68	14200	0.0263
15.36	14100	0.0147
30.72	14200	0.0331
61.44	14100	0.0339
122.88	13800	0.0307

to this thesis and predominantly focus on reducing the NMSE. As a result of the above works altering the RLS algorithm authors, such as [78] and [79] , achieve an improved NMSE at the cost of increased computational complexity.

Each model order of non-linearity encounters the onset of instability at a different time sample for the same dataset. The calculation of minimum error is estimated using equation (4.8), for different memory lengths depending on the nonlinear order of the model as shown in equation (4.4). Therefore there is a finite input length of training signal $\boldsymbol{X}(n)$ that can be used for a model before instability occurs, regardless of the value of non-linear order, memory tap length and sampling frequency, as can be seen comparing Tables 4.1, 4.2 and 4.3.

The influence of increasing the number of memory taps is much greater than that caused by increasing the order of nonlinearity of the model or altering the sampling frequency [81]. The effect of increasing the memory taps alone can be seen in Table 4.1. The effect of increasing the nonlinear order of the model can be seen in Table 4.3.

The Tables 4.1, 4.2 and 4.3 utilise a wideband code division multiple access (WCDMA)

Table 4.3: Training signal length relationship to model order of nonlinearity

Nonlinear order	Maximum signal length(samples)
2	14100
3	12900
5	6600
7	6400

5MHz bandwidth single carrier signal sent through a Doherty PA at 2.6GHz. For the experiment the input signal is set in order to drive the PA into saturation and both the input and output are sampled at 30.72 MHz. Figure 4.1 presents the transmitted and received signals by the Volterra model used in this investigation, given by equation (4.1).

$$y(n) = \sum_{p=1}^{P} y_p(n) \tag{4.1}$$

Where,

$$y_p(n) = \sum_{i_1=0}^{N-1} \cdots \sum_{i_P=0}^{N-1} h_p(i_1, \cdots, i_p) \prod_{i=1}^{p} x(n - i_r)$$

Where $x(n)$ and $y(n)$ is the input and output signal to the system respectively. $h_p(i_1, ..., ip)$ represents the filter co-efficient expansion utilising, p, the highest order for the non-linearity of the Volterra series expansion. N represents the maximum memory tap length chosen [34].

4.2 The RLS training algorithm

Recursive Least Squares (RLS) is a well known training method used for training the behavioural models of PAs [66, 67]. RLS is an iterative form of least mean squares [46] that converges to a minimum error while training a model after fewer iterations than LS [66, 67].

The following equations give a mathematical description of the RLS algorithm to minimise the cost function in equation (3.24) by minimising the error ϵ to update $H(n)$ and the update matrix $\boldsymbol{C}^{-1}(n)$ in an iterative fashion, heuristically as in the

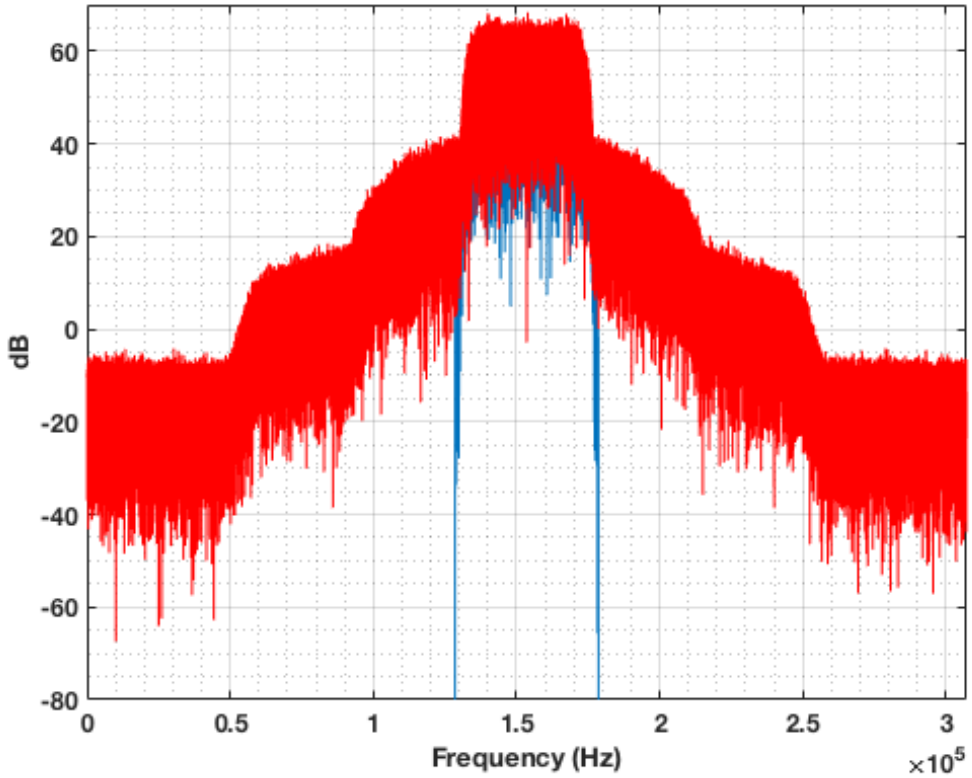

Figure 4.1: Input and output signals sent through Doherty PA

conventional RLS algorithm [46] where $K(n)$ depicts the gain vector.

$$\epsilon(n) = d(n) - \boldsymbol{X}^T(n)H(n-1) \tag{4.2}$$

$$K(n) = \boldsymbol{C}^{-1}(n)\boldsymbol{X}(n) \tag{4.3}$$

$$\boldsymbol{C}^{-1}(n) = \frac{\boldsymbol{C}^{-1}(n-1)}{\lambda} - \frac{K(n)\boldsymbol{X}^T(n)\boldsymbol{C}^{-1}(n-1)}{\lambda} \tag{4.4}$$

$$H(n) = H(n-1) + K(n)\epsilon(n) \tag{4.5}$$

Where,

$$\boldsymbol{C}(n) = \sum_{k=1}^{n} \lambda^{(n-k)}\boldsymbol{X}(k)\boldsymbol{X}^{\boldsymbol{T}}(k) \tag{4.6}$$

$$e(n) = d(n) - \boldsymbol{X}^T(n)H(n) \tag{4.7}$$

$\boldsymbol{C}(n)$ depicts the weighted least squares auto-correlation function, $\boldsymbol{X}(k)\boldsymbol{X}^T(k)$, of the $1 \times N$ dimensional input vector $\boldsymbol{X}(n)$.

When updating the RLS algorithm model coefficients as seen in equation (4.5), the size of $\boldsymbol{C}^{-1}(n)$ is determined by the total number of a given models coefficients given a particular memory length and model order of nonlinearity. Increasing the order of nonlinearity and memory tap length of the model increases the size of $\boldsymbol{C}^{-1}(n)$, thus eigen decomposition can become extremely complex.

One objective when performing DPD and behavioural modeling is to identify the most computationally efficient structure which can accurately characterise the behaviour of a PA, enabling the calculation of the most concise set of coefficients to predistort a signal successfully.

Small increments in either the order of nonlinearity or memory depth can greatly increase the number of model coefficients, which in turn can severely reduce the computational efficiency of the model. However, the opposite scenario where the nonlinear order or memory depth can be reduced will greatly reduce the computational effort required.

4.3 Auto correlation function

Equation (4.4) of the RLS training algorithm presents that the expected value of $\boldsymbol{C}^{-1}(n)$ is a function of the auto correlation matrix. When examining an isolated sample of the input signal, $\boldsymbol{X}_n$, it is treated as a random variable with expected values in the form of $E\{\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}}\}$. This is used in equation (4.6) and expanded in equation (4.8) to illustrate the behavior of the calculation .

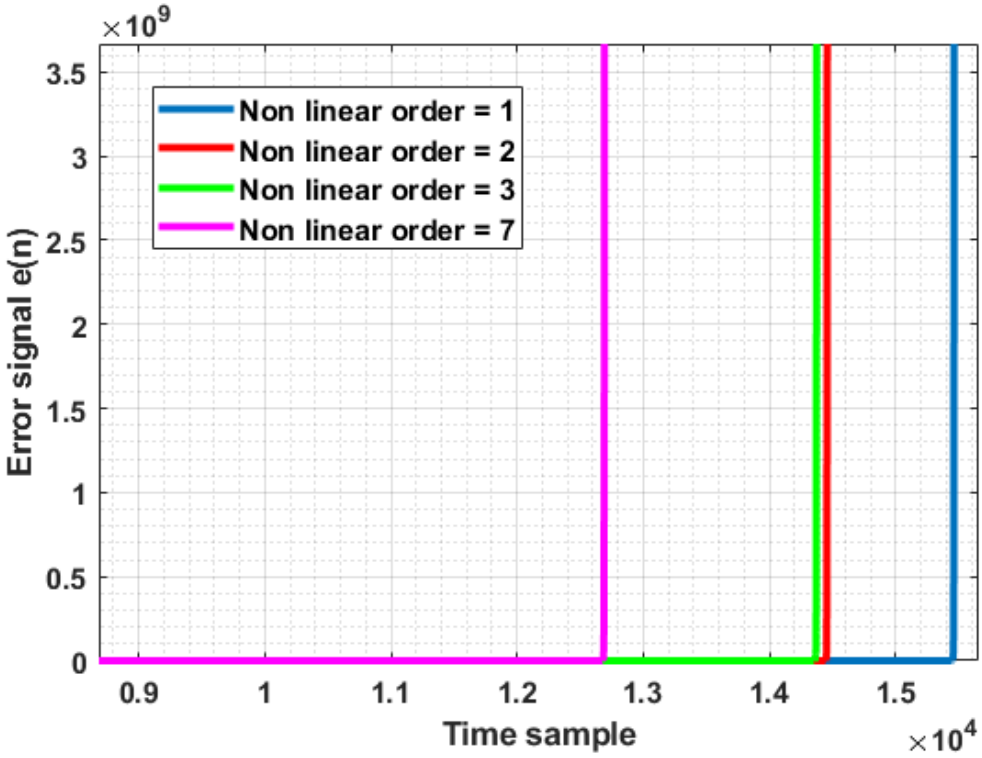

Figure 4.2: An illustration of the error signal increasing in magnitude versus time samples of the input training signal. Instability is indicated by a rapid increase in magnitude, as shown for each nonlinear order of the Volterra model.

$$\boldsymbol{R_{xx}} = \begin{bmatrix} E[\boldsymbol{X}_1\boldsymbol{X}_1] & E[\boldsymbol{X}_1\boldsymbol{X}_2] & \dots & E[\boldsymbol{X}_1\boldsymbol{X}_n] \\ E[\boldsymbol{X}_2\boldsymbol{X}_1] & \dots & & \vdots \\ \vdots & & & \\ E[\boldsymbol{X}_n\boldsymbol{X}_1] & E[\boldsymbol{X}_n\boldsymbol{X}_2] & \dots & E[\boldsymbol{X}_n\boldsymbol{X}_n] \end{bmatrix} \tag{4.8}$$

Equation (4.8) assumes that all of the components are real random vectors. Should the vectors be considered as complex values random vectors $\boldsymbol{R_{xx}}$ must be in Hermitian form [46], which is not realisable in every model when using various memory and nonlinear order values.

The autocorrelation function contains eigenvectors. Eigenvectors are vectors that have both direction and magnitude. The direction of the eigenvectors tend towards the point of convergence [82], between two data points, or sampled points. Once the point of convergence, or minimum error, is reached the eigenvectors and values become oscillatory, continuing to attempt to point in the direction of largest variance

[83].

With a memory length of N, the auto-correlation matrix takes into account N previous values of $x(n)$. Using a large value for N memory taps leads to an accelerated convergence of the minimum error. Aggregating the new values of $\boldsymbol{C}(n)$ causes incorrect values of the eigenvalues, which represent the variance of the data along the eigenvector directions. This is because equation (4.8) includes memory taps which have introduced variance and been calculated as dissimilar to the input signal $X(n)$.

4.4 A novel early stopping criterion

RLS is an iterative form of the LS estimation. For a linear system, the LS estimate is given by

$$Ax = b \tag{4.9}$$

The LS solution calculates a value of x such that Ax is the closest value as possible to b. LS exploits the fact that $\|b - Ax\|$ is the square root of a sum of squares. Consider A to be an $n \times m$ matrix, b is in R^m, the LS solution of (4.9) is a value of x in R^n such that

$$\|b - A\hat{x}\| \leq \|b - Ax\| \tag{4.10}$$

for all x in R^n. Considering (4.10) graphically in vector form the LS estimate deems that Ax will be in column space A ($C(A)$), as it is inherently limited to $C(A)$. LS calculates a value of x such that Ax is as geometrically close to b as possible in terms of distance, such that $\|Projection(b) - b\| \to 0$

$$A\hat{x} - b \in C_A \tag{4.11}$$

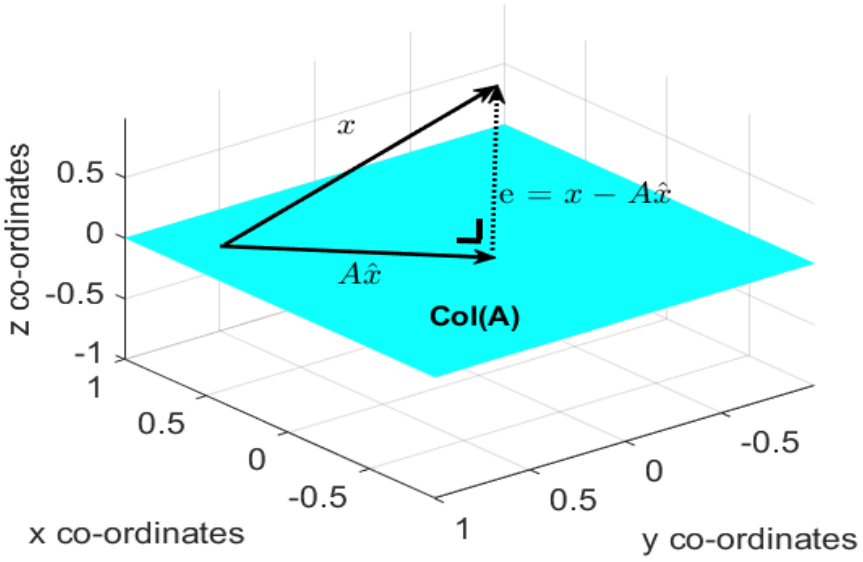

Figure 4.3: Illustration of orthogonality of LS estimation on column space A

$\hat{b}$ must be the orthogonal projection of b on to $Col(A)$, for the solution of $Ax = \hat{b}$ to be valid. This entails that $Ax = \hat{b}$ is consistent and that there is a solution of $\hat{x}$ in R^n. By the orthogonal decomposition principle, the projection has the property that $b - \hat{b}$ is orthogonal to $Col(A)$ [84].

In RLS the cost function as seen in (3.24) can be written as equation (4.12).

$$J(n) = \sum_{k=1}^{n} \lambda^{(n-k)}(d(k) - \hat{d}(n))^2 \tag{4.12}$$

As illustrated above in Figure 4.3 $\hat{d}(n)$ must be orthogonal to $d(n)$, a change in phase indicates a change in the $\hat{d}(n)$ projection. $\lambda^{(n-k)}$ is inherently a scalar value. Multiplying by $\lambda^{(n-k)}$ will not change the phase of $\hat{d}$ as the imaginary term will always be 0.

Aforementioned related works focused on improving the NMSE or adding additional computational complexity by altering the RLS algorithm, thus increasing latency. This chapter determines a method by which to cease training of the model using RLS before instability occurs. Instability can be circumvented by a simple observation of a value that is pre-existing natively in the conventional RLS algorithm and therefore

does not increase computational complexity, minimising latency.

The update matrix as in equation (4.4) is calculated by the difference of two separate matrix manipulations, with one matrix manipulation containing the more current information of the autocorrelation function, consisting of the right hand side of equation (4.4). Henceforth, this will be referred to as the change in update matrix and denoted it as $\triangle C^{-1}(n)$, as defined in equation (4.13). Equation (4.4) is a difference equation which has the potential to generate eigenvalues that result in a divergence from the trajectory of minimum error [46]. To avoid the estimated output of the model diverging from the least squares error, previous authors have examined eigen analysis of the auto correlation function. This involves altering the limits specified in the auto correlation function based on the statistical analysis of the specific input training signal and, therefore, requiring individual computations for each respective training signal.

$$\triangle C^{-1}(n) = \frac{K(n)\boldsymbol{X}^T(n)\boldsymbol{C}^{-1}(n-1)}{\lambda} \tag{4.13}$$

As RLS minimises the linear least cost function, the phase $\triangle C^{-1}(n)$, given by equation (4.14) , is expected to tend toward the projection of the minimum error with consistent phase, i.e the desired output and actual output are tending toward the same point in the complex plane. Significant deviations in the complex values of $\triangle C^{-1}(n)$ from the original trajectory indicates definitively the onset of instability. The phase of $\triangle C^{-1}(n)$ was chosen as the stopping criterion as it does not add any additional computational complexity as it can be carried out as an observation.

$$\theta(\triangle C^{-1}(n)) = tan^{-1}\frac{Imaginary(\triangle C^{-1}(n))}{Real(\triangle C^{-1}(n))} \tag{4.14}$$

By observing $\triangle C^{-1}(n)$ on a sample per sample basis it is possible to identify the sample point at which instability begins to occur. As the auto correlation function relates $X(n)$ and $X(n-1)$, which contain $N-1$ common elements, and therefore

should remain highly similar to previous values. Observing the first element of the matrix $\triangle C^{-1}(n)$, allows for a comprehensive measurement of the eigen vector behaviour as the diagonal values of $\triangle C^{-1}(n)$ will be identical as seen in equation (4.8). In this way, a deviation in the sample to sample values in $\triangle C^{-1}(n)$ indicates a deviation from the trajectory towards the minimum error.

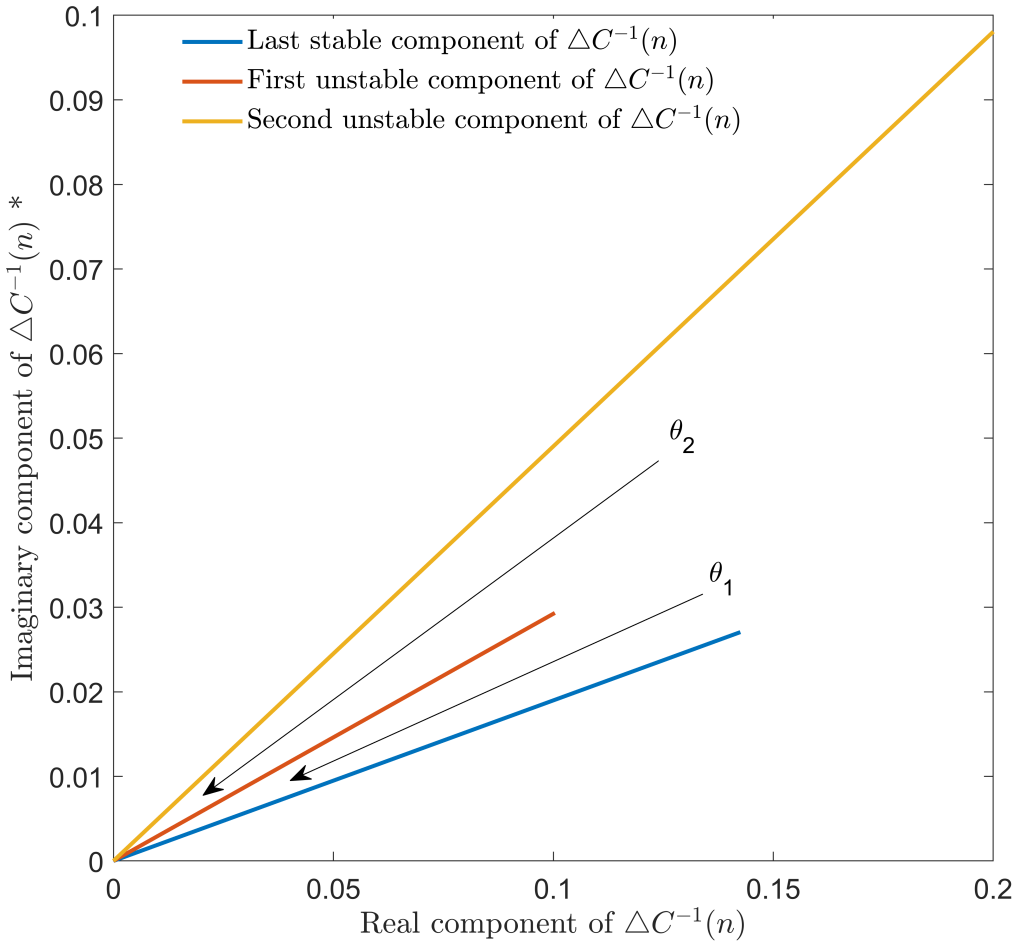

Figure 4.4: Illustration of phase discrepancy, of $\triangle C^{-1}(n)$. θ_1 depicts the phase of the first element of $\triangle C^{-1}(n)$ is 0.1876 Radians. θ_2 is 0.5278 Radians. * The limits of this figure have been truncated for aesthetic purposes. Please note the magnitude of the second unstable $\triangle C^{-1}(n)$ extends to co-ordinates $0.5891 + 0.2888i$

Figure 4.4 depicts how stability may be inferred from the change of phase for the first element in $\triangle C^{-1}(n)$. As previously stated, stability is indicated by the plotted vectors remaining in close proximity to the trajectory of minimum error. A change in direction and sudden increase of the magnitude of the vectors indicates that the

estimate $\hat{d}(n)$, is tending away from the plane of the least squares estimation of the error.

While a simulated PA will not introduce any external errors into the model, experimental validation may incur errors such as those resulting from noise. As such it is necessary to introduce a threshold into the early stopping criterion to prevent premature termination of training. The tolerance we suggest is that the phase component of the first element of $\triangle C^{-1}(n)$ should fall between ∓ 0.25 Radians. The phase of the first element of $\triangle C^{-1}(n)$ rises rapidly, as can be seen in Figures 4.7 (a) and 4.4. Therefore ∓ 0.25 Radians was considered to be a suitable threshold for ceasing training prior to an extreme divergence of the phase component. This allows for the RLS algorithm to preserve high fidelity of the estimated output without major alteration to the RLS algorithm.

In Figure 4.4 it can be seen that the model behaviour has become imbalanced i.e. the vectors have exceeded ∓ 0.25 Radians prescribed tolerance. Once the point of convergence, or minimum error, is exceeded the eigen vectors are becoming oscillatory and increase in magnitude, i.e. attempting to point in the direction of largest variance, ∞ [83]. Therefore, it is beneficial to cease training once the point of convergence has been exceeded as defined by the prescribed threshold that is applied to the resulting surrogate measure obtained from the first value of the matrix given by equation (4.8).

4.5 Experimental validation

In order to validate the early stopping criteria proposed in this work a variety of single carrier signals are sent from an AD-FMCOMMS3 evaluation board, through a Doherty PA RFHIC RTP26010N1 driven by a driver amplifier (NXP BGA7210) centered at 2.6GHz. The corresponding input and output signals are sampled at 30.72 MHz.

As can be seen in Figure 4.5 the hardware configuration to complete model extrac-

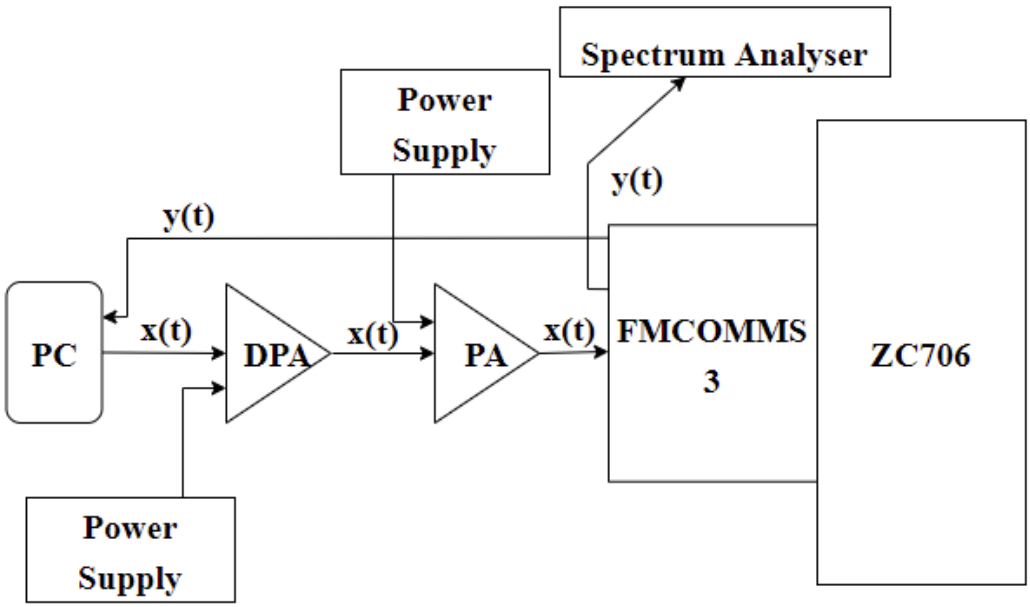

Figure 4.5: Experimental hardware setup

tion was a combination of the ZC706 development board and FMCOMMS3. The SMA cables were connected from the a transmitter to the input of the PA (NXP BGA7210), into the RFHIC high power PA and connected back via a reciever port from the output of the PA. A spectrum analyser, Rhode & Schwarz FSL, was used to visualise nonlinearities.

Various degrees of nonlinearity were measured by the proposed setup. Signal power was increased to induce severe saturation as shown in Figure 4.6. Prompting this response from the PA enabled the authors to discern the fidelity of the PA model and ensure robust modelling.

To illustrate the onset of instability an arbitrary memory length and order of nonlinearity was tested. For the purposes of illustration both values were set to 2 producing Figure 4.7. Each of the four modulation schemes become unstable at different time samples (n) as seen in 4.7 (a). DVBS2x becomes unstable at n= 1.3742×10^4, WCDMA becomes unstable at n=1.385×10^4, 5G NR becomes unstable at n =1.425×10^4 and OFDM LTE becomes unstable at n = 1.428×10^4.

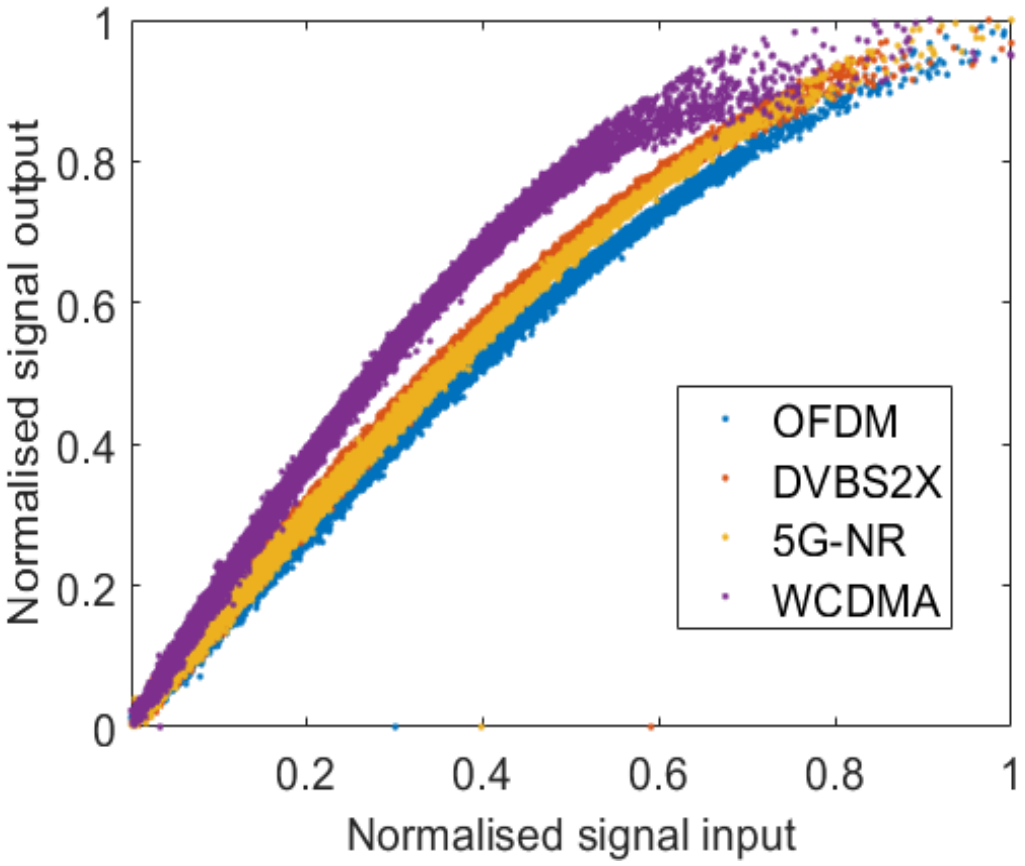

Figure 4.6: AMAM plot of examples of input output signal pairs sent and received through the PA using AD-FMCOMMS3

Figure 4.7 (b), plots the experimentally measured output of the PA versus the estimated output of the PA model utilising the proposed algorithm. The early stopping criterion algorithm ceased training prior to instability as defined by the phase contained by the first complex value of $\triangle C^{-1}(n) > 0.25$ Radians. Both model estimated outputs were compared to the experimentally measured output in terms of the Normalised Mean Square Error (NMSE) as illustrated by Table 4.4.

Figure 4.8 illustrates the onset of instability with regard to the phase, imaginary and real component of $\triangle C^{-1}(n)$. The imaginary and phase components of the first element of $\triangle C^{-1}(n)$ diverge by a large amount close to the point of onset of instability.

The model utilising the proposed algorithm returned acceptable model accuracy in terms of NMSE values, as seen in Table 4.4. The NMSE values listed in the table

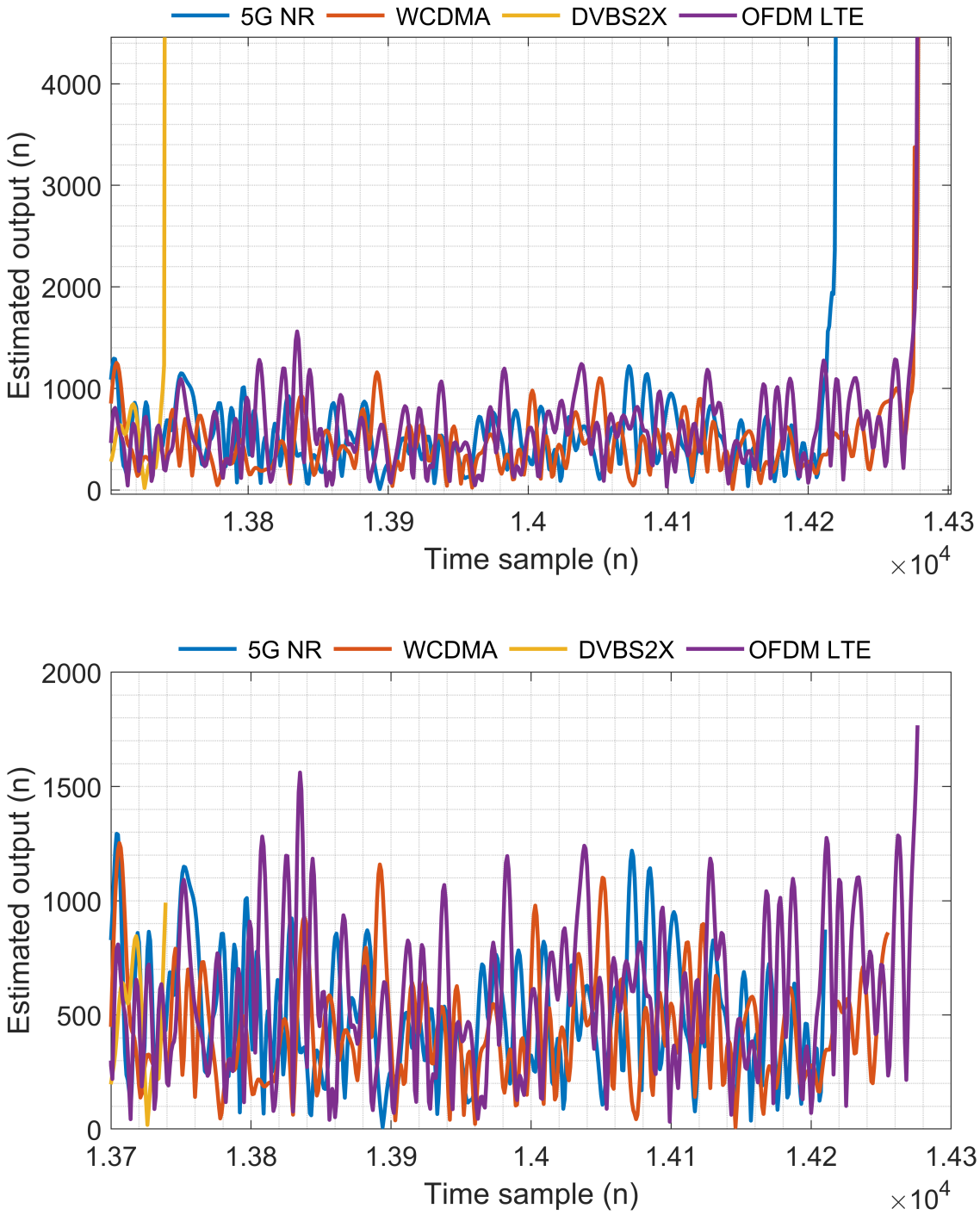

Figure 4.7: Experimentally measured output signal (a) estimated signal output without early stopping criterion and (b) with proposed early stopping criterion. It can be seen in (b) that the input training signal length has been truncated prior to the onset of instability. Let it be noted that for illustrative purposes only samples from 13700 onwards are depicted

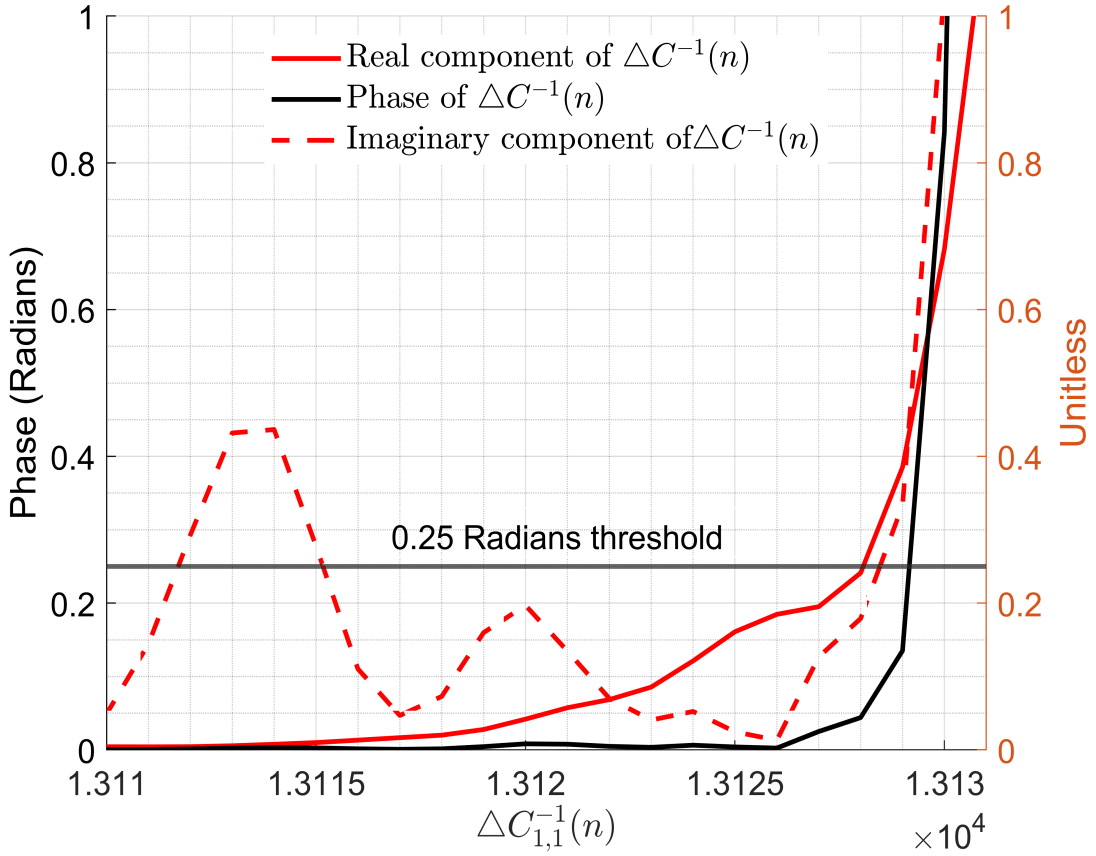

Figure 4.8: Real, Imaginary and Phase components of the change in the first element of the update matrix, $\triangle C^{-1}(n)$, as the early stopping criterion is surpassed. The early stopping criterion is shown as a constant 0.25 Radians threshold that indicates instability when exceeded by the phase. Phase was chosen rather than the real or imaginary components as the indicating factor in order to maximise the input training signal length. Let it be noted the limits of this Figure have been truncated for illustrative purposes. This Figure was produced using a nonlinear model order and a memory tap length of 3

show that, by utilising the early stopping criterion, the NMSE value indicates high fidelity between the estimated output and the actual output (visually illustrated in Figure 4.9). Severe degradation of the NMSE values occurs rapidly after this point, as shown by the NMSE values listed for +10 and +20 samples after the early stopping criterion recommends the cessation of training. Not only will the early stopping criteria prevent the training routine producing unstable outputs, but it will also maximise the length of the input training signal. Therefore allowing the continued reduction of the model coefficient error, maximising the accuracy of the extracted model.

Table 4.4: A comparison of NMSE values when stopped using early stopping criterion(ESC), 10 samples beyond ESC, 20 samples beyond ESC

Signal Standard	**ESC**	**ESC+10**	**ESC + 20**
WCDMA	**-25.514 dB**	**-24.737 dB**	**5.286 dB**
DVBS2X	**-25.427 dB**	**-14.459 dB**	**14.941 dB**
LTE OFDM	**-24.904 dB**	**-24.903 dB**	**4.118 dB**
5G-NR	**-24.4079 dB**	**-23.615 dB**	**-3.136 dB**

Let it be noted that this experiment was conducted with various nonlinear orders and memory lengths. The early stopping criteria operated as expected, ceasing training before the onset of instability, for all values of memory lengths and nonlinear orders. As expected from Figure 4.2, the input training signal length approached instability earlier as the memory lengths and nonlinear orders were increased.

Figure 4.9, plots the experimentally measured output of the model trained with distinct signal standards versus the estimated output of the model using the proposed algorithm. All signals were 5MHz bandwidth single carrier signals sent through a Doherty PA at 2.6GHz, the same experimental procedure as mentioned above.

The NMSE values shown in Table 4.4 for the various signal standards that the PA model is accurate up to the point where the early stopping criterion identifies when to cease training. Three of four signals NMSE values dis-improve marginally up to ten time samples after the early stopping criterion is met. However, twenty time samples after the early stopping criterion is met, the extracted model NMSE values have dramatically degraded in all four cases. The NMSE values in Table 4.4 indicate that, through the application of the proposed early stopping criterion, training is terminated prior to instability occurring. This does not require alteration of the RLS algorithm, only a monitoring of a single value that is inherent to the calculation.

Further experimentation was carried out in order to investigate the proposed early stopping criterion. The same experimental set up was used to that above to qualify the early stopping metric across a range of input power levels and bandwidths. The PA used was a 2.6GHz Doherty PA and thus a 10MHz OFDM signal was chosen for further experimentation.

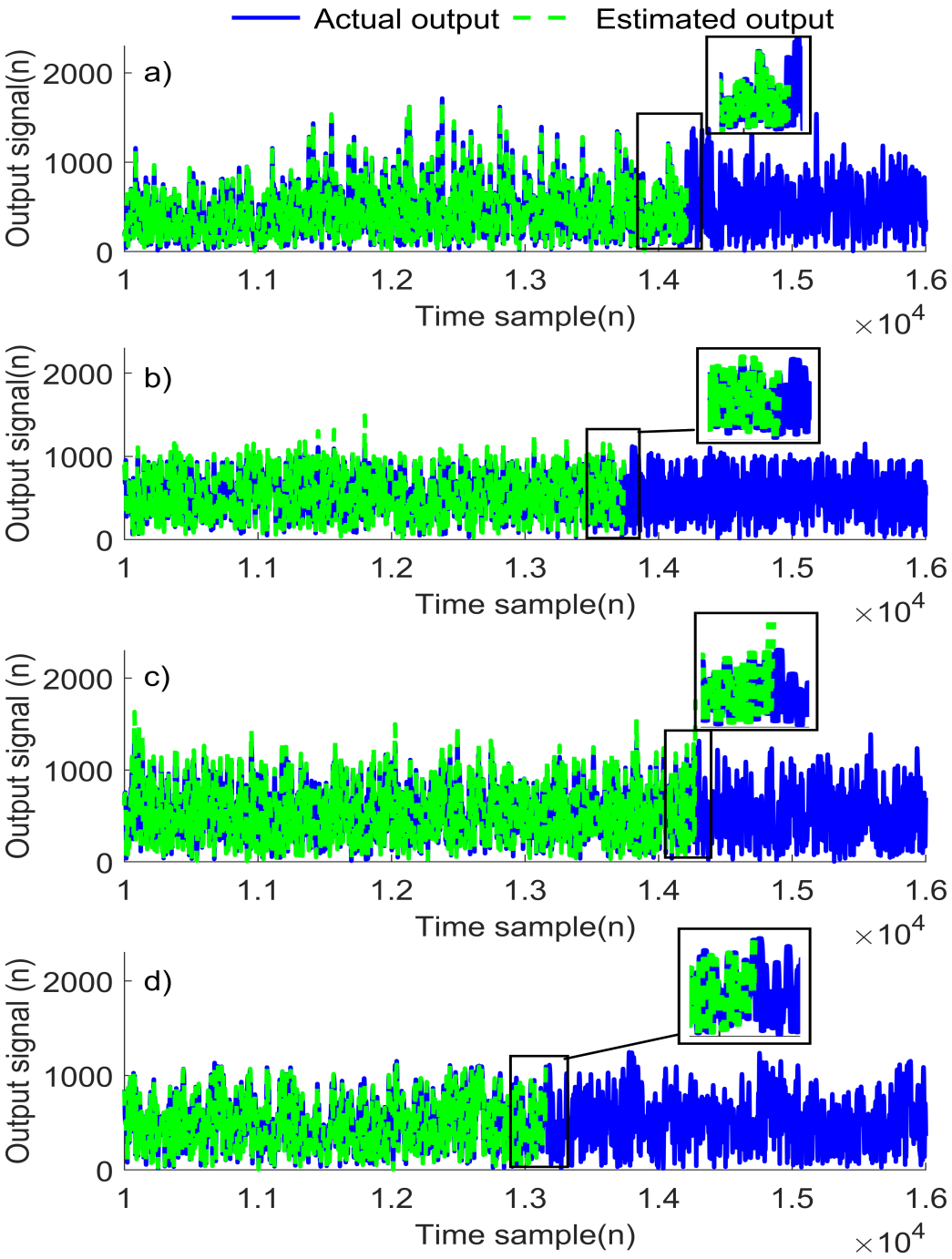

Figure 4.9: Experimental signal output versus estimated signal output for various signal standards with proposed algorithm (a) 5G-NR, (b)DVBS2X, (c)LTE OFDM and (d) WCDMA. In each of these cases it can be seen that the estimated output corresponds closely with the experimentally validated output, indicating that the applied algorithm does not negatively affect the modeling capabilities. Zoomed in sections have been provided for clarity. Let it be noted that for illustrative purposes only samples from 10^3 onwards are depicted

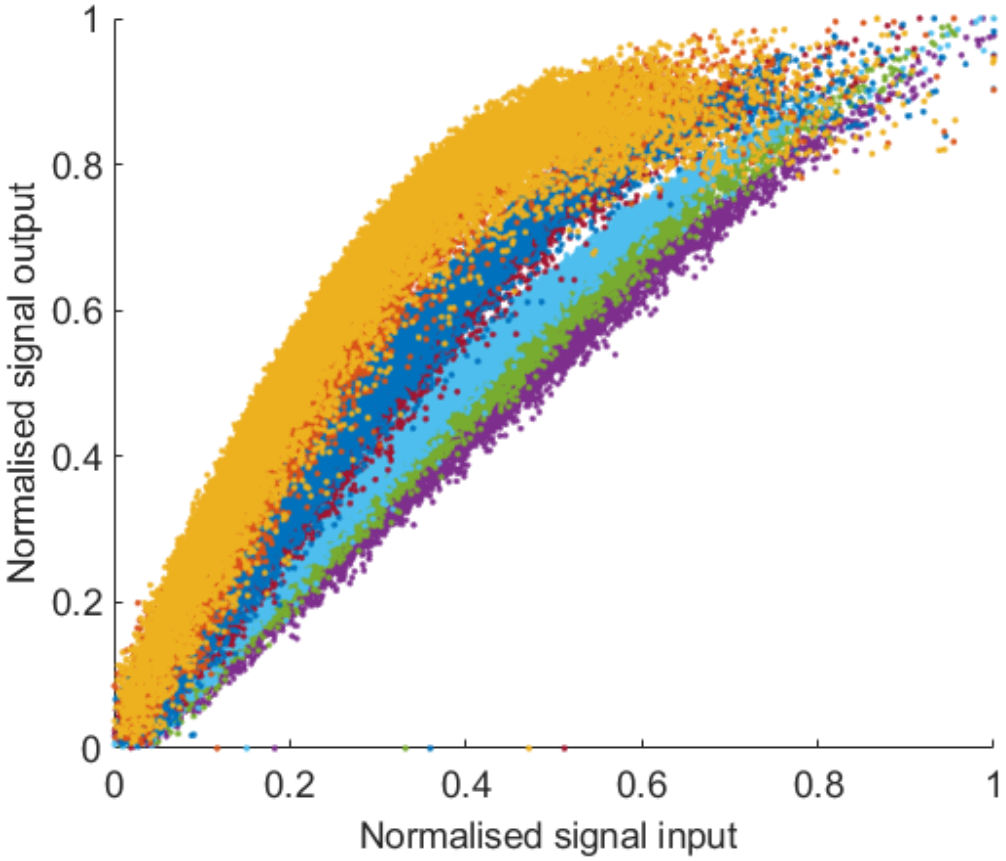

Figure 4.10: AMAM of multiple 10MHz OFDM signals transmitted through a PA at different input powers to depict the nonlinearity elicited and validated using proposed novel ESC. Memory effects of the input output signal pairs is much more severe than that in Figure 4.6

Table 4.5 qualitatively confirms that the proposed early stopping criterion operates as expected for 10MHz bandwidths. It should be noted that the early stopping criterion in this case reduces the number of input training sample length more prematurely than that of the 5MHz experiment. This is due to the memory effects present in the 10MHz input output signal pairs that was not as prevalent in the 5MHz AMAM plot as seen in Figure 4.6, when compared to that of Figure 4.10. By placing a threshold on the phase of the first element of $\triangle C^{-1}(n)$, limits the phase between the possible values between all estimations of $\hat{d}(n)$, a wider AMAM curve contains values of $\hat{d}(n)$ that are more distant from one another. The below Table 4.5 was determined using the original phase threshold of 5 degrees, in order to extend the input training signal length the threshold can be set higher.

Table 4.5: 10MHz OFDM experiment

Input Power of signal (dBm)	**NMSE at ESC (dB)**	**ESC+20**	**ESC+50**
-18.34 n=13980	-55.8743	-55.8894	40.4820
-17.34 n=13512	-55.3910	-55.4080	30.5491
-16.34 n=13627	-55.5664	-55.5832	22.1272
-15.34 n=13342	-55.2836	-55.3009	-9.8311
-14.34 n=14038	-55.6723	-54.8082	419.3612
-13.34 n=13780	-55.5562	-55.5855	168.2840
-12.34 n=13812	-55.3174	-55.3492	170.1731
-11.34 n=13653	-55.2963	-55.3317	126.0777
-10.34 n= 13700	-55.1963	-55.2074	57.4600
-4.34 n=13604	-55.0205	-28.6387	159.183
-6.34 n=13028	-54.0881	-54.0973	42.3758

4.6 Chapter Summary

In conclusion, this chapter provides an early stopping criterion to identify the onset of instability of a polynomial model during RLS training. Experimental validation of the proposed procedure shows that the NMSE of the experimental output vs estimated output indicates high fidelity until the point identified by the early stopping criterion is exceeded after which point it deteriorates rapidly. NMSE values are detailed in Table 4.4 and visually illustrated by Figures 4.4, 4.7, and 4.8. Application of this early stopping procedure eliminates the need to apply supplementary computational analysis, thereby minimising the computational complexity required to guarantee a stable model while maintaining it's accuracy. Automating this early stopping procedure is conveniently implemented, only requiring the observation of

a pre-existing value that is produced by the RLS algorithm, the autocorrelation based update. This method has been experimentally validated for a high power 10 Watt PA using four distinct radio access techniques namely LTE OFDM, WCDMA, DVBS2X and 5G-NR.

In the proceeding chapter a novel DPD technique will be introduced that allows for reduced computational complexity.

Chapter 5

DPD Dimension Reduction

As described in Chapter 1, complex configurations of PA architectures such as MIMO and beamforming are commonly utilised to improve the overall system performance, by increasing throughput of data [2]. The complexity of DPD solutions has grown considerably in recent years as the performance demands of cellular network communications become ever more challenging. PA characteristics may change depending on input stimulus. Therefore with arrays of PAs experiencing continuous changes in operating conditions, ideally DPD parameters must also be continuously adapted.

A DPD solution should ideally produce a highly linear PA output signal using the minimum number of coefficients. Overfitting and ill-conditioning is an area of concern when training DPD coefficients [85]. Feature selection and/or extraction has been a popular research topic concerning the reduction of DPD function dimension which inherently mitigates against overfitting.

5.1 Related works

DPD coefficient estimation is typically performed using an iterative optimisation algorithm such as the LS, as described in Chapter 4. Typically DPD is over determined [85]. Polynomial-based DPD exhibits structural multicollinearity between predictors, enabling researchers to intelligently prune DPD coefficients that do not

contribute to the efficacy of the DPD linearisation [86, 87]. Multi-collinearity refers to coefficients adding little to no value to the DPD structure and will be explained further within this chapter.

Authors of [88] allow for function reduction and a change of basis by employing Principal Component Analysis (PCA). Authors of [89] use a partial least squares (PLS) algorithm, which allows for the basis matrix used in the DPD estimation to be transformed at every iteration. The Orthogonal Matching Pursuit (OMP) algorithm is used to determine which basis functions contributed most to the DPD adaption. The drawback of these aforementioned methods is the complex sorting algorithms necessary in order to rank effectiveness of each coefficient on their respective models, resulting in high computational costs.

In this chapter, a method is introduced for selectively partitioning the DPD coefficient updates. Basis functions within a DPD model can be selectively partitioned, and updated separately. The rationale of updating only a portion of the DPD coefficients is that all coefficients when calculated do not converge uniformly over training iterations [90].

The Frisch–Waugh–Lovell (FWL) theorem [91] can be applied to allow a local update of a single partition while the remaining coefficients are held constant. The FWL theorem gives a formula for partitioned LS estimates and shows that residuals from any multiple of sequential regressions are identical [91].

5.1.1 Multicollinearity

Modeling and DPD inaccuracies of Volterra based models typically a result of poor parameter estimation. Poor parameter estimation referring to incorrect order of nonlinearity chosen, inaccurate memory length and also statistical features of basis functions. Basis functions are all possible values of input signal permutations multiplied by model coefficients to replicate the output signal of a PA. There are inherent structural multicollinearity between basis functions. The inherent multicollinearity

is due to each coefficient being determined using the same signal - i.e. more than two basis functions used during parameter estimation are highly linearly related, where there is very little or no additional information given by the preceeding basis function.

Multicollinearity is problematic in DPD/ behavioural modeling as it increases computational cost while adding very little value to the parameter estimation.

Consider the least squares batch estimation technique described in Chapter 3 and given by equation (5.1).

$$h = (Y^T Y^{-1}) Y^T e \tag{5.1}$$

As described in Chapter 3, h refers to model coefficients, deduced from a data matrix of basis functions Y, solving the LS estimation problem. Should coefficient values of Y repeat there is no improvement on the projection to the minimum error.

5.2 Methods commonly used to mitigate against multicollinearity

Feature selection refers to the process in which redundant parameters are removed from a larger set of DPD parameters. Although these operations are extremely useful the algorithms needed to sort which parameters are valuable and/or redundant are typically computationally expensive.

5.2.1 Partial Least Squares (PLS)

Authors of [89] perform PLS inside a DPD adaptation loop to actively adjust the basis function matrix in a subsystem. The basis reduction is implemented iteratively allowing for a dynamic DPD loop that operates according to the residual linearisation error, defined as the difference between the actual and the desired linear PA

output signals. PLS regression is a technique that reduces the basis functions to a smaller set of uncorrelated components and performs least squares regression on these components, instead of on the original data, unlike LS. Authors of [89] use a computationally costly algorithm, the Dynamic OTM algorithm is used to reduce the number of predictors into a smaller data set. As discussed above, this extra computation results in longer latency but yields excellent results.

5.2.2 Principal Component Analysis (PCA)

PCA has been experimentally validated by previous authors for use in feature selection in DPD [92]. PCA performs eigen-decomposition using Singular Value Decomposition (SVD), where the LS estimation process sorts the diagonal elements of the SVD into values aligned in decreasing order of magnitude.

The reduced SVD of the matrix A can be expressed as (5.2) [93], where redundant columns and rows have been omitted.

$$A = U\Sigma V^T. \tag{5.2}$$

Where U is an n by n unitary matrix, Σ is an n by m diagonal matrix and V is an m by m unitary matrix.

The first element in the matrix corresponds to the element that contributes the most to the estimated output signal, and the last the least contribution. Authors of [92] truncate the matrix at a point in which the remaining matrix diagonal coefficients describe the majority of the variance between parameters, i.e. there is a sudden drop in magnitude across the coefficients. PCA calculates the point in which it is determined that a particular singular value is the last singular value of relatively high contribution. Truncating the SVD matrix at this point will eliminate the features that account for the lowest variance, leaving the remaining terms unaffected. The reduced set of DPD coefficients h can now be calculated as (5.3).

$$h = V_p \Sigma^{-1} U^T e. \tag{5.3}$$

Where the diagonal elements of the reduced matrix Σ coefficients describe the majority of the variance between parameters.

5.2.3 Regularisation

Regularisation is a technique where a function adds a penalty or cost to an optimisation function [94]. Ridge regression is a technique of regularisation that adds a bias, λ to the LS coefficient estimation process. The LS expression can now be expressed as equation (5.4) [94].

$$\widehat{h} = (Y^T Y + \lambda I)^{-1} Y^T e \tag{5.4}$$

Where I is the identity matrix. The process is implemented such that as λ increases, the magnitudes of the estimated coefficients shrink. The minimisation problem to obtain $\hat{h}$, shown in (3.25), is adjusted to obtain the best fit for a biased set of DPD coefficients, $\hat{h_\lambda}$, as seen in equation (5.5).

$$\hat{h_\lambda} = \arg\min_h \sum_{i=1}^{n} |u_i - \sum_{j=1}^{m} Y_{ij} h_j|^2 + \lambda \sum Y_{ij} = u - Yh^2 + \lambda Y \tag{5.5}$$

Ridge regression-based techniques are inherently susceptible to the effects of outliers [94].

5.2.4 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is a regularisation technique performed by the addition of the absolute value of the coefficients as a bias to the LS loss function [95], and seen in equation (5.6).

$$h = (Y'Y)^{-1}Y'e - \frac{\lambda_1}{2}b \tag{5.6}$$

Where b is ± 1 depending on the sign of the corresponding coefficient and λ_1 is the shrinkage coefficient. LASSO is a regularisation technique that has been previously experimentally validate both PA models and PA DPD coefficients [96–98].

5.3 The Frisch Waugh Lovell Theorem (FWL)

In order to fully understand the Frisch Waugh Lovell (FWL) Theorem there are two critical types of matrices used to manipulate the linear algebra proof of FWL. The first essential matrix type is the projection matrix. A projection matrix (P) is a linear function on a vector space v, such that when its order is increased it remains equal i.e. $P^2 = P^5$. Considering a linear projection matrix, p, on to v, the vector space which is a two-dimensional line. The projection of x is equal to $P\bar{x}$, where Px returns the point on v that is 'closest' to x. This can be seen graphically in Figure 5.1, where the green line shows the orthogonal projection and the red lines show non orthogonal projections.

'Closest' typically implies euclidian distance given between two points p and q by equation (5.7).

$$d(p,q) = \sqrt{\sum_{i=1}^{N}(p_i - q_i)^2} \tag{5.7}$$

Where N refers to the maximum dimension of the vector space.

In matrix form the projection matrix can be derived as follows. As above a point x in N dimensional space, vector space v and project x on to v such that the minimisation is given by equation (5.8).

$$argmin_c = \sqrt{\sum_i (x_i - x)^2} = argmin_c \sum_i (x_i - x)^2 = argmin_c \sum_i (cv_i - x)^2 \tag{5.8}$$

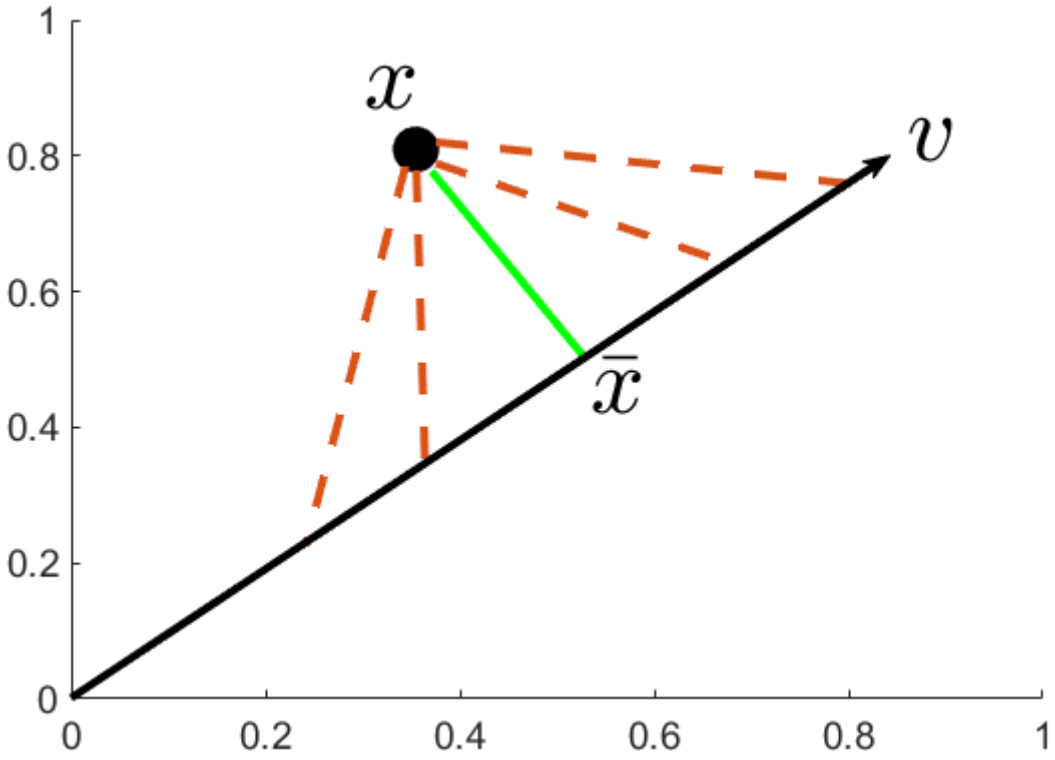

Figure 5.1: Graphical representation of linear projection

This rearrangement can be done as the square root is a monotonic transformation. The optimal value of c is equal as any potential point x along v is a scalar multiplication of the line cv. Equation (5.8) is then differentiated resulting in equation (5.9).

$$\frac{d}{dc}\sum_i (cv_i - x)^2 = \sum_i 2v_i(cv_i - x)^2 = 2(\sum_i cv_i^2 - \sum_i v_i x) \tag{5.9}$$

Converting equation (5.9) from summation notation to matrix multiplication results in equation (5.10).

$$2(\sum_i cv_i^2 - \sum_i v_i x) = 2(cv'v - v'x) \implies 0 \tag{5.10}$$

Solving equation 5.10 gives equations (5.11).

$$
\begin{aligned}
2(cv'v - v'x) &= 0 \\
cv'v - v'x &= 0 \\
cv'v &= v'x \\
c &= (v'v)^{-1}v'x
\end{aligned}
\tag{5.11}
$$

Hence the projection of x on to v gives equation (5.12).

$$P = v(v'v)^{-1}v' \tag{5.12}$$

In terms of regression notation, vector coefficients of linear regression are typically expressed using X, such that equation (5.12) can be written as equation (5.13).

$$P = (X'X)^{-1}X'X \tag{5.13}$$

The projection matrix when applied to an outcome vector y produces a set of predicted values. Hence the relationship between linear projection and regression can be seen in equation (5.14).

$$\hat{y} = (X'X)^{-1}X'Y \equiv P_xY \tag{5.14}$$

The second matrix needed is the residual maker matrix M. M is the compliment of P, as seen in equation (5.15).

$$M = I - (X'X)^{-1}X'X \tag{5.15}$$

The matrix M can now be manipulated as seen in equation (5.16).

$$My = y - (X'X)^{-1}X'y \equiv y - Py \equiv \epsilon \tag{5.16}$$

Where ϵ is the residuals from regressing Y onto X. The complementary relationship between M and P can now be deduced as in equations (5.17).

$$\begin{aligned} y &= \bar{y} + \epsilon \equiv Py + My \\ Iy &= Py + My \\ Iy &= (P + M)y \\ I &= P + M \end{aligned} \tag{5.17}$$

By leveraging the relationship of the matrices above FWL states the basis function Y as seen in (5.18), can be segmented in to two or more sections such that

$$Y = Y_1 h_1 + Y_2 h_2. \tag{5.18}$$

Where Y_1 and Y_2 are the partitioned basis functions of size $n \times k_1$ and $n \times k_2$. The partitioned sets of coefficients h_1 and h_2 are regression coefficients respectively. FWL asserts that it is possible to re-specify a linear regression model by manipulating the residual of only one partition of the basis function. The equation (5.18) can be re-specified according to the FWL theorem as

$$\hat{u}M_1 = M_1 Y_1 \hat{h_1} + M_1 Y_2 \hat{h_2} + M_1 e. \tag{5.19}$$

$\hat{u}$ is the estimated output, e is the error between the estimated and actual output. M_1 is the residual maker of Y_1, which encapsulates the variation of Y_1 that cannot be resolved by Y_2 and given by (5.20). $M_1 Y_1 \hat{h_1} = 0$, as the regression of Y_1 on itself yields no variance unexplained by Y_1

$$M_2 = I - Y_2 (Y_2^H Y_2)^{-1} Y_2^H. \tag{5.20}$$

The superscript H denotes the Hermitian transpose. A partitioned LS regression

can now be performed. To do this, (5.23) can now be solved using FWL such that

$$\hat{h} = (Y_1^H M_2 Y_1)^{-1} Y_1^H M_2 e. \tag{5.21}$$

In this way only the coefficients from the first or targeted partition are updated. The second partition remains static. LS is a global estimator, employing all coefficients regardless of convergence rate in its estimation as seen in (5.22). Allowing for reduced computational complexity, i.e LS must perform a matrix inversion costing ($O(C^2N)$), versus FWL ($O(C^3)$). DPD coefficients do not converge globally. Partitioning Y supports select coefficients to be updated locally, supporting convergence of the partitioned basis functions, in turn directly influencing the trajectory of the global error more rapidly.

For this chapter the indirect learning adaption was used and is defined as the minimisation problem (5.22) and seen in Figure 3.15.

$$\hat{h} = ||u - Yh||^2. \tag{5.22}$$

The matrix Y is an m by n matrix of signal permutations of the PA output signal, y, equivalent in selection to those in X.

The DPD coefficients were extracted by solving (5.23).

$$\hat{h} = (Y^H Y)^{-1} Y^H e. \tag{5.23}$$

The error signal, e, for an ILA is calculated as (5.24) [99].

$$e = u - \hat{u}. \tag{5.24}$$

Where the post-distorted signal, $\hat{u}$, can be expressed as (5.25).

$$\hat{u} = Y\hat{h}. \tag{5.25}$$

5.4 Experimental Results

To validate the proposed technique an experimental testbench was developed, as shown in Figure 5.2. The testbench consisted of a Skyworks SKY66297-11 PA, an Analog Devices AD9375 transceiver board, MATLAB on a local PC and a Rohde & Schwarz FSL spectrum analyser. A 40 MHz 5G-NR signal was passed into the PA and the output signal captured using the observation receiver of the AD9375. The PA output was also attenuated and monitored on the spectrum analyser to confirm correct operation of the PA.

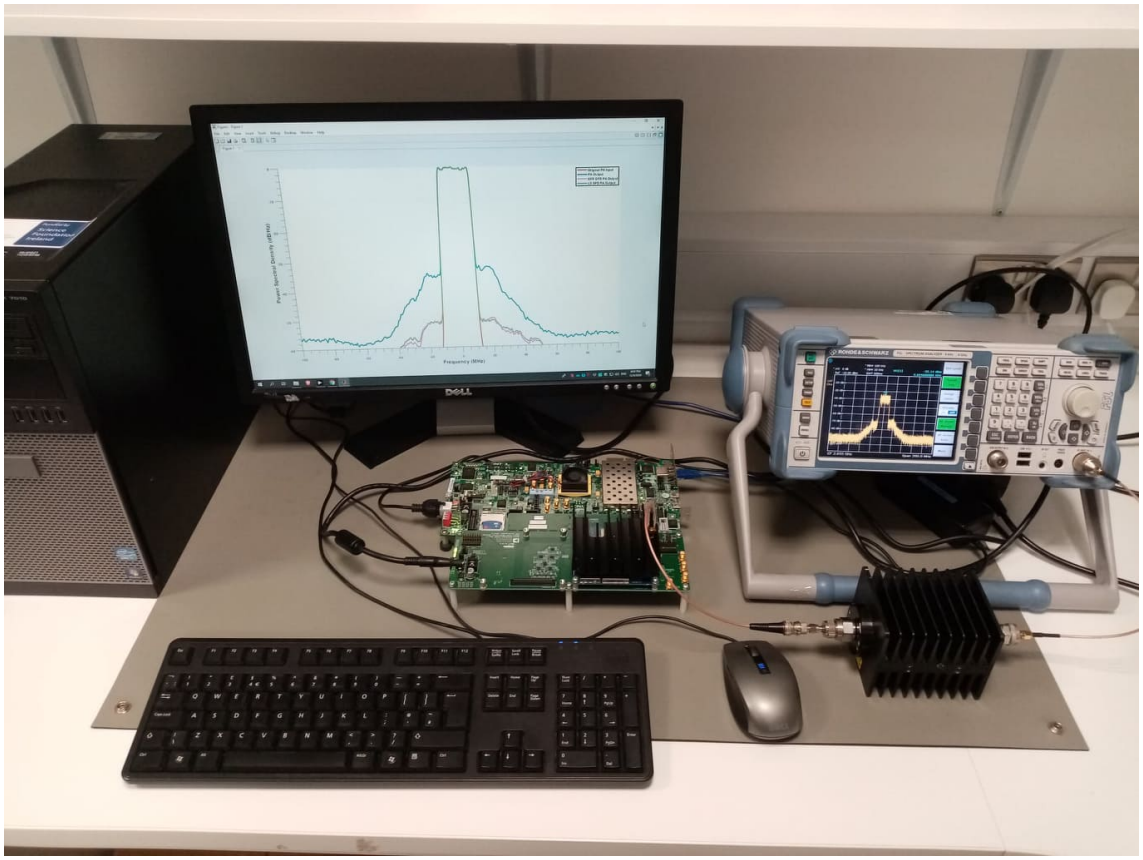

Figure 5.2: Experimental hardware setup for experimentation on novel FWL DPD technique

As a reference, a 103 coefficient GMP based DPD function was used to linearise the

PA. The LS and LS with FWL methods were both used to train the DPD coefficients and subsequently compared. The LS and FWL method achieved a superior normalised mean square error (NMSE) performance.

The relative performance of the two approaches over two iterations can be seen in Table 5.1. The NMSE, error vector magnitude (EVM) and number of coefficients updated for both the LS and LS with FWL methods are displayed in Table 5.1.

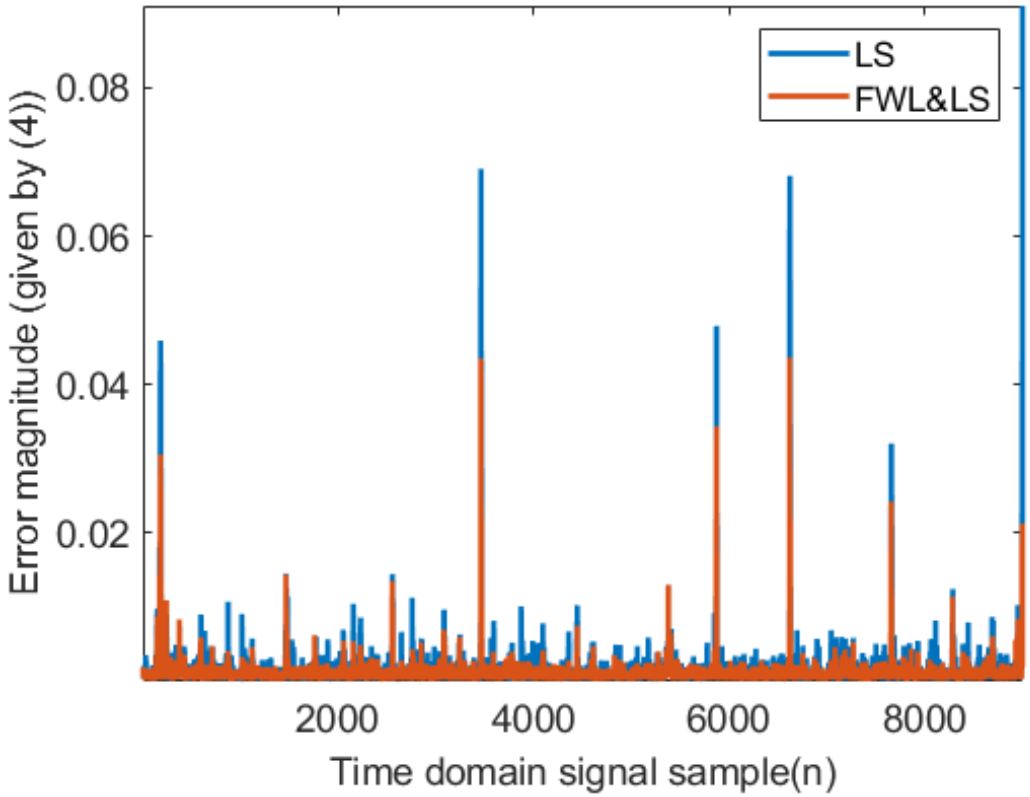

Figure 5.3: Error signal for LS and FWL with LS

Table 5.1: Performance Comparison

Method	LS	FWL
Coefficients Updated	206	107
NMSE (dB)	-40.4597	-44.8611
EVM (%)	.9484	.5714

The experimental results in Table 5.1 show that updating only a select partition of the basis functions can more effectively linearise the dynamic nonlinear PA distortions. By applying LS with FWL to train the DPD coefficients allows for convergence

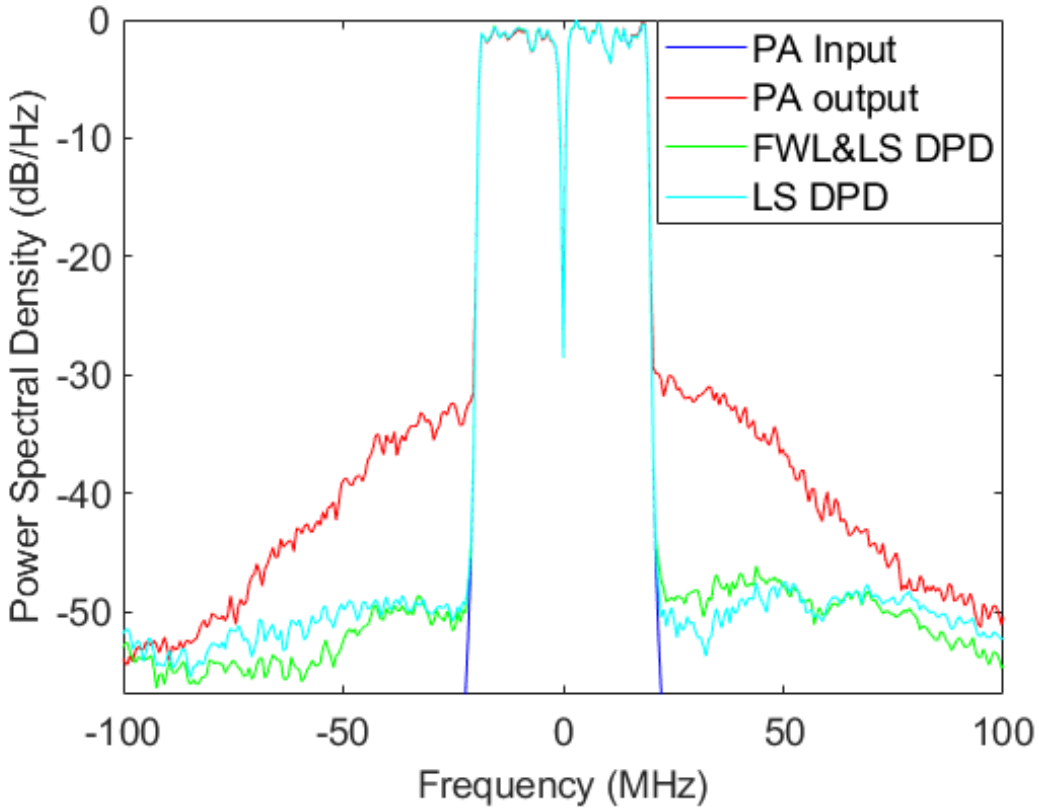

Figure 5.4: Frequency domain representation at baseband

of the partitioned basis functions to achieve a lower error. This is evidenced in the results by the decreased error value of the LS with FWL in Figure 5.3.

Figures 5.4 and 5.5 demonstrate the linearisation and experimental validation constructed using LS with FWL and LS linearisation techniques on a 40MHz 5G-NR signal. Additional computational complexity due to pruning algorithms on the model are avoided. A frequency domain plot of the linearised PA output, and the original PA input and output is shown in Figure 5.4. AMAM characteristics for the successful pre-distortion of FWL LS are also presented in Figure 5.5.

5.5 Conclusions

This chapter demonstrates a method for selective partitioned adaptation for a LS trained DPD system. The novel methodology presented exploited the Frisch-Waugh-Lovell technique to enable a more accurate adaptation. The partitioned regression

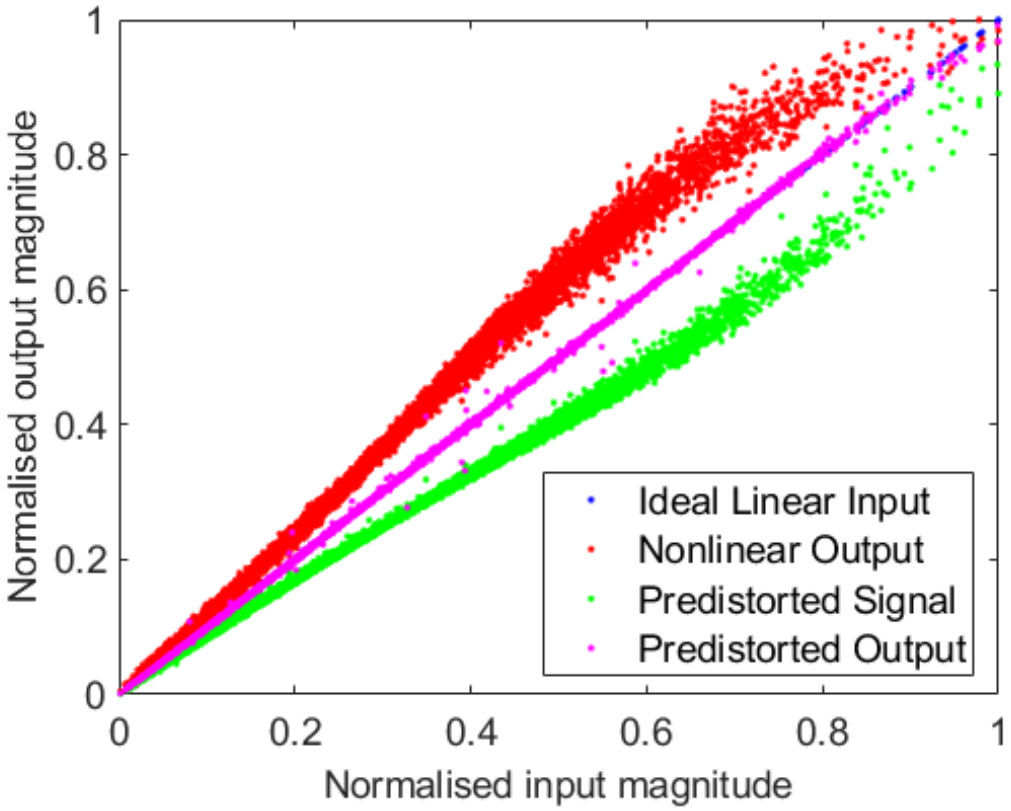

Figure 5.5: AM/AM plot

was accomplished by applying the FWL theorem to a DPD function trained using LS regression. The proposed methodology was experimentally validated by adapting a particular subset of a GMP DPD function which linearised a PA amplifying a 40MHz 5G-NR signal.

The partitioning of model parameters is a research question unto itself. Partitioning specific effects introduced by complex PA architecture may be adapted in further work to allow for partitioning and updating segments of specific basis functions.

In the next chapter a novel DPD and modelling metric is introduced that allows for reduced computational complexity and latency.

Chapter 6

Metric for training signal dimension reduction for high frequency power amplifier modeling and DPD

This chapter details a method to determine the maximum dimension reduction that can be applied to the PA input signal while maintaining accurate modeling or pre-distortion. A novel metric is presented that can identify if the PA input signal is of sufficient characterisation to determine model coefficients for PA behavioural modeling or DPD.

This chapter implements downsampling as an example of a pre-processing method as it is an attractive solution of dimension reduction to implement in both DPD and behavioural modeling. Although pre-processing by downsampling has been proven using arbitrary downsampling factors, there is no previously published metric to determine whether or not a downsampling factor will provide accurate DPD or modeling coefficients.

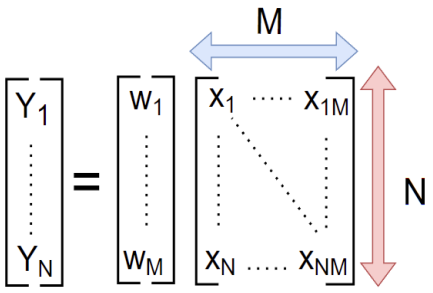

Figure 6.1: Matrix illustration of equation (6.1)

6.1 Related works

Downsampling is an attractive option to lessen the computations during model and DPD coefficient extraction of PAs. Coefficients are typically determined using the LS equation as seen in equation (4.9), but can be written more generically by equation (6.1).

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} \tag{6.1}$$

Where $\mathbf{y}$ is the $N \times 1$ PA output, $\mathbf{w}$ is the $M \times 1$ vector of parameters and $\mathbf{X}$ is the $N \times M$ data matrix, with N being the number of samples and M being the number of basis functions of the model. The structure of this is shown in Figure 6.1.

The aim of feature selection is to reduce the size of M in the matrix $\mathbf{X}$ of equation (6.1), as discussed in Chapter 5. Downsampling allows us to reduce N, notably $N \gg M$. However reducing N has been done in the following related works in an ad-hoc manner.

Reducing the sampling rate when computing PA DPD coefficients has recently been an active area of research [100, 101], identifying that relaxed sampling requirements can be utilised when performing DPD.

Reducing the cost of hardware implementation for DPD can be achieved by sampling at a lower rate or 'critically sampling' the PA output signal. Using a low sampling rate on the output signal to model or pre-distort these systems yields a dataset that

facilitates a computationally simpler solution.

Authors of [102] use the GMP model in conjunction with LS error correction to design an algorithm. The algorithm combines a minimal number of probabilistic based samples to approximate an autocorrelation function to match histogram data of consecutive samples. Work produced by [102] requires less computations than standard LS GMP model, while maintaining the accuracy of the traditional consecutive behavioural modeling method.

Recent previous works by [103] utilised time interleaved DPD to achieve a reduction in sampling rate to 1.5 of the signal bandwidth. The proposed methodology presents that the introduction of a sliced multi-stage DPD can achieve comparable results to that of conventional DPD. [103] method introduces latency due to the introduction of supplementary time delays and increased computational complexity. Authors of [104] propose a model in which a band-limiting function is incorporated in the calculated DPD coefficients, while achieving comparable sampling rates as authors of [103]. Bandlimited DPD requires an additional band limiting filter that may be difficult and costly to design and implement [105].

Authors of [106] suggest the use of a reduced sampled Transmitter Observation Receiver (TOR) output at an intermediate frequency. The experimental results in the case of a ′non flat receiver′ requires additional computational complexity to mitigate against non ideal effects of the frequency response of the custom built TOR, without being considered degrades the accuracy of the DPD to a greater extent than without DPD implemented. Work by [106] resulted in achieving a sub-sampling factor of 100 consisting of a data set of 80 thousand samples.

6.2 Theoretical background

The probability density function (PDF) of a modulated signal, for both I and Q samples is Gaussian, as the signal is treated as a random variable. Overlaying the I and Q waveform probability curve in x- and y-axis illustrates that the IQ coordinates

are proportional to the power of the signal, as depicted in Figure 6.2.

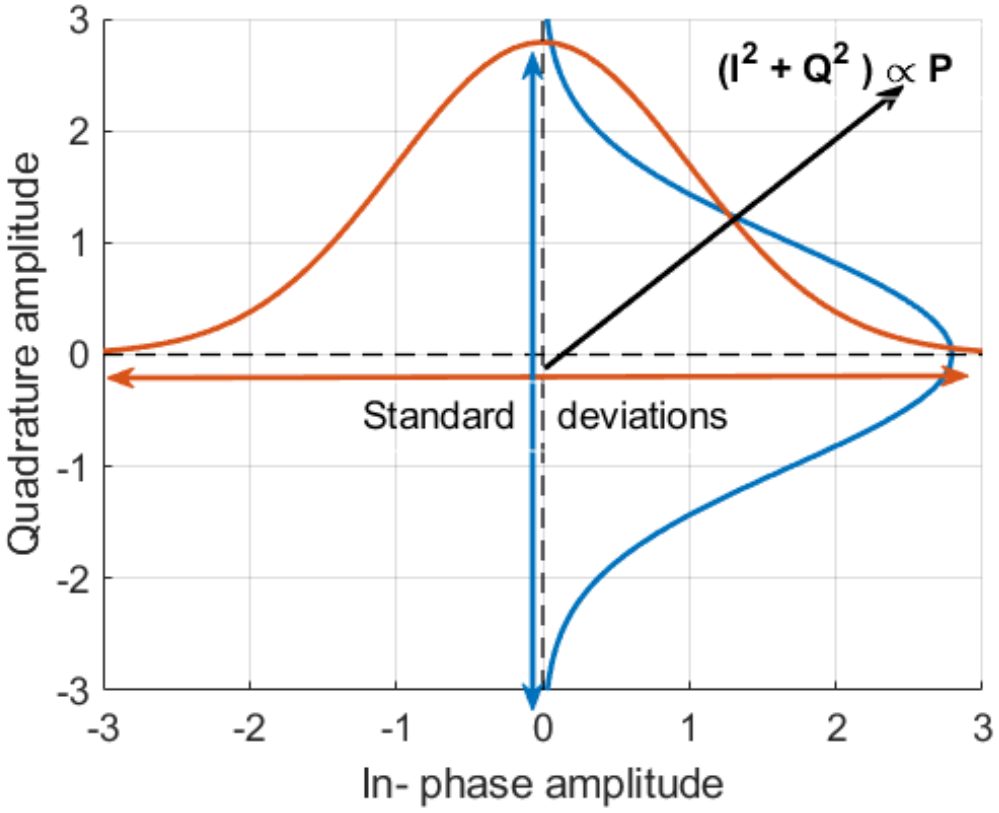

Figure 6.2: Illustration of Gaussian PDF of I and Q voltage waveforms, depicting the maximum standard deviation and relationship to power

Information signals undergo modulation techniques based on the signal standard being implemented. Consider, as in Figure 6.2, the QPSK modulation technique. The distance between the real to real and imaginary to imaginary modulation points depict the maximal standard deviation measure of voltage probability of the real or imaginary components. The complex values of the modulated signal are proportional to the resulting power level of the signal, indicating that maximum standard deviation captures maximum power levels of the signal.

The central limit theorem states that as independent random variables are sampled more frequently their distribution tends towards a normal distribution, which is depicted by the PDF of the independent random variable, or signal sample [46]. The normal PDF is given by equation (6.2).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \tag{6.2}$$

Where μ, σ and x refer to the mean, standard deviation and input signal values respectively.

The above equations indicate that the PDF of an input signal alters with mean and variance. The modulation of a signal limits the variance that a signal will undergo to it's maximum standard deviation in the digital domain.

The PDF of a random variable is completely defined once the mean and standard deviation are specified [107]. Modulated signal variance is inherently limited, as shown in Figure 6.2. As depicted by equation (6.2) the characteristics of the signal's PDF are determined by the mean and standard deviation of the signal, entailing that if these values are maintained the signal may be represented by a lower number of sample points.

The expected value ($E(x)$) of a PDF is calculated by equation (6.3) and can be seen to contain an exponential indicating that there is a sharp increase or decrease in the function dependant on the values of the mean and standard deviation. In other words equation (6.3) is sensitive to any changes in mean or variance.

$$E(x) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} x e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = \mu \tag{6.3}$$

6.2.1 The complementary cumulative distribution function

The complementary cumulative distribution function (CCDF) curve is used to determine the PAPR of an information signal. This is done as in Figure 6.3 [108] by following the steps below.

1. A modulated waveform varies voltage with time. The waveform shows the voltage of either I or Q in the time domain as in Figure 6.2.

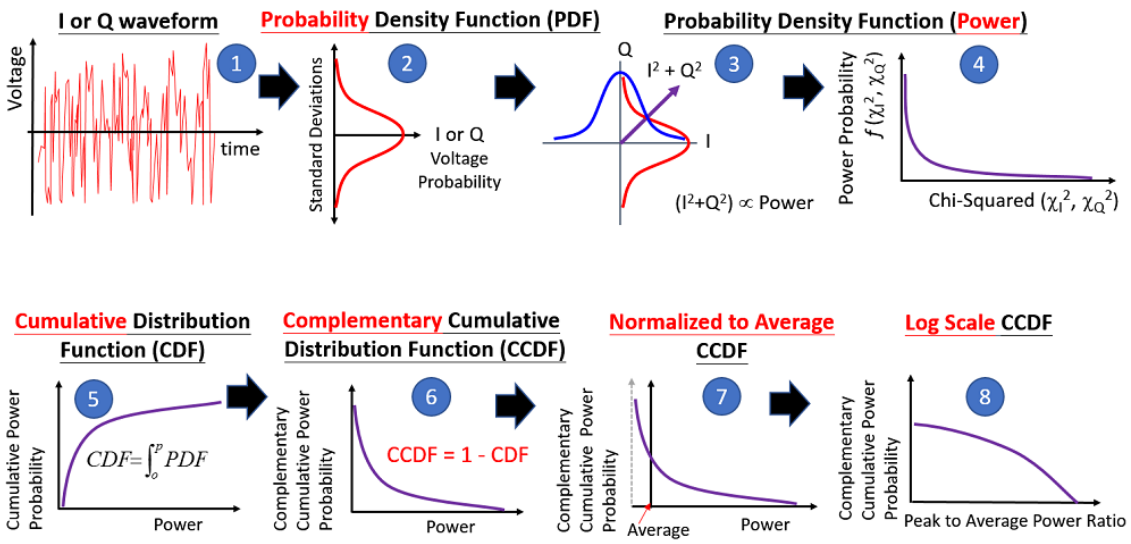

Figure 6.3: The construction of a CCDF Curve

2. The I or Q waveform is then converted into a probability density function curve. The probability of a modulated signal for I or Q is a Gaussian PDF.

3. The I and Q waveforms are overlayed on the probability curve in both the x and y axes.

4. The power of the waveform is proportional to the sum of the square of I and square of Q. The summation of the square of two normal distributions is a Chi-squared distribution with two degrees of freedom. The curve is called power PDF.

5. By integrating the PDF curve, it can be shown that the probability of the power is below the upper integration limit. This integrated function is called the Cumulative Distribution Function (CDF).

6. By subtracting the CDF from 100% probability and obtaining the "Complementary" CDF ($CCDF = 1 - CDF$), which indicates how likely the power is to be at or exceed a given level.

7. The function is then normalised to determine the average power.

8. The log scale is used to illustrate more resolution of signal levels that occur at very low probability, in other words infrequently.

6.2.2 Confidence intervals

A confidence interval displays the probability that a parameter will fall between a pair of values around the mean. The selection of a confidence level for an interval determines the probability that the confidence interval produced will contain the true parameter value i.e. the mean value. Table 6.1 lists the confidence interval for a Gaussian PDF.

Table 6.1: Confidence intervals of Gaussian PDF

Confidence Interval	**Probability (%)**
$\mu + \sigma$	**68**
$\mu + 2\sigma$	**95**
$\mu + 3\sigma$	**99.7**

The standard deviation is a measure of the amount of variation of a set of values. A low value of standard deviation indicates that the values tend to be close to the mean. A high standard deviation indicates that the values are dispersed out over a wider range.

For a random variable A consisting of $i \rightarrow N$ data points, the standard deviation (σ) is defined as in equation (6.4).

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}|A_i - \mu|^2} \tag{6.4}$$

Where μ is the mean of A.

Gaussian data forms a bell shape when plotted on a graph, with the sample mean in the centre and the remaining data distributed on either side of the mean.

The confidence interval (CI) for data which follows a standard normal distribution with sample mean $\overline{x}$ is given in equation (6.5).

$$CI = \overline{x} \pm Z^* \frac{\sigma}{\sqrt{N}} \tag{6.5}$$

Z^*, the Z score, is a measure of how many standard deviations away from the mean must be encapsulated in order to reach the desired confidence level, given by equation (6.6).

$$Z^* = \frac{\overline{x} - \mu}{\sigma} \tag{6.6}$$

- $Z^* > 0$ indicates that the x-value is greater than the mean.
- $Z^* < 0$ indicates that the x-value is less than the mean.
- $Z^* = 0$ indicates that the x-value is equal to the mean.

6.2.3 Digital sub sampling

The nature of the transceiver is that a signal is modulated pre-transmission according to the modulation technique chosen. The time domain signal, pre-modulation, is considered to be constructed of random variables. The modulated signal co-ordinates, or power values, can be treated as standardized random variables, the resulting PDF of both I and Q waveforms are Gaussian, as depicted by Figure 6.2, implying that the maximum standard deviation of the waveform gives the highest power peaks, as the co-ordinates furthest from the mean, located at the extremities of the PDF.

It is widely accepted to use consecutive sampling data sets when training a behavioural model or computing DPD coefficients. In the samples the highest peak of the time domain signal should be included. Multiple training sample selection strategies have been proposed by previous works, such as consecutive sampling encompassing the highest peak in a data set, random sampling, sampling in order to reconstruct the behaviour of a PDF and subsampling. Subsampling, by previous works as discussed in the theoretical background, exhibit that subsampling to a high degree is possible.

The mathematical expression proposed by this chapter, as given in equation (6.7),

offers a method to determine if a downsampled signal can be used for accurate modelling or pre-distortion, using simple statistical values.

6.3 Novel metric

$$\frac{\mu}{\sigma^2} \approx \frac{\mu_{OS}}{\sigma^2_{OS}} \tag{6.7}$$

$\frac{\mu}{\sigma^2}$ is a coefficient of variation, a measure of dispersion of data around the mean, and is the metric proposed to be used when performing DPD and PA modeling. Please note the subscript OS refers to original signal. As described in the above theory this metric will indicate the need to cease pre-processing of a signal. $\frac{\mu}{\sigma^2}$ was chosen as it combines the effects of both equations (6.3) and (6.4).

As will be discussed further in this chapter, the effect of equation (6.4) with large data sets is that of a scalar multiplier of $\frac{1}{N-1}$. As the N scalar value reduces, the reliance of the mean and variance increases. This is because there are $N-1$ degrees of freedom, which refers to the maximum number of logically independent values, which are values that have the freedom to vary, in the data sample set. An increase in the degree of freedom allows for a better fit to data, because more freedom is allowed in the model structure. Thus with a low value of N, an alteration in mean or variance will effect the coefficient extraction greatly.

Equation (6.7) cannot account for mean bias. Should the mean alter greatly due to pre-processing there is an alternate metric given as equation (6.8).

$$\frac{\Delta\mu}{\sigma^2} = 1 \tag{6.8}$$

Where $\Delta\mu$ refers to the difference in mean between the original input signal and pre-processed signal. σ^2 refers to the variance of the pre-processed signal. This equation entails that the mean and variance should ideally increase proportionally. An increase in both the mean and variance would entail that although the mean is

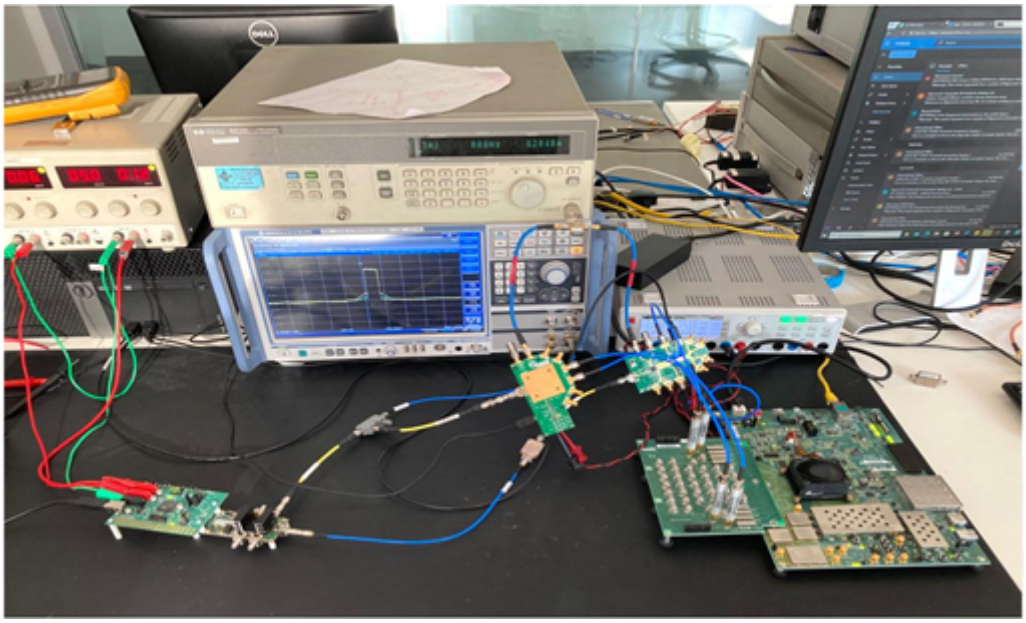

Figure 6.4: ZCU111 DPD testbench

altering, the variance is altering in proportion to account for this change.

6.4 Experimental validation : DPD

To validate the proposed technique an experimental mmWave testbench was developed. A 40MHz 5G NR signal was transmitted using an RFSOC ZCU111. The signal was then transmitted through an NXP PA centred at $26GHz$. The experimental set up can be seen in Figure 6.4. A GMP model, as described in previous chapters, was used to determine the DPD coefficients. The same number of DPD iterations, 10 , was done using the same number of GMP coefficients, 53, in order to produce a fair comparison. The linearisation routine used in all experiments was batch LS, as discussed in Chapter 3.

The 40MHz 5G NR input signal was downsampled only when the DPD coefficients were being calculated, thus reducing the computation of the inverse DPD matrix as discussed in previous chapters. The DPD coefficients were then applied to the original input signal. Successive downsampling by a factor of two was used as this is the most computationally efficient solution.

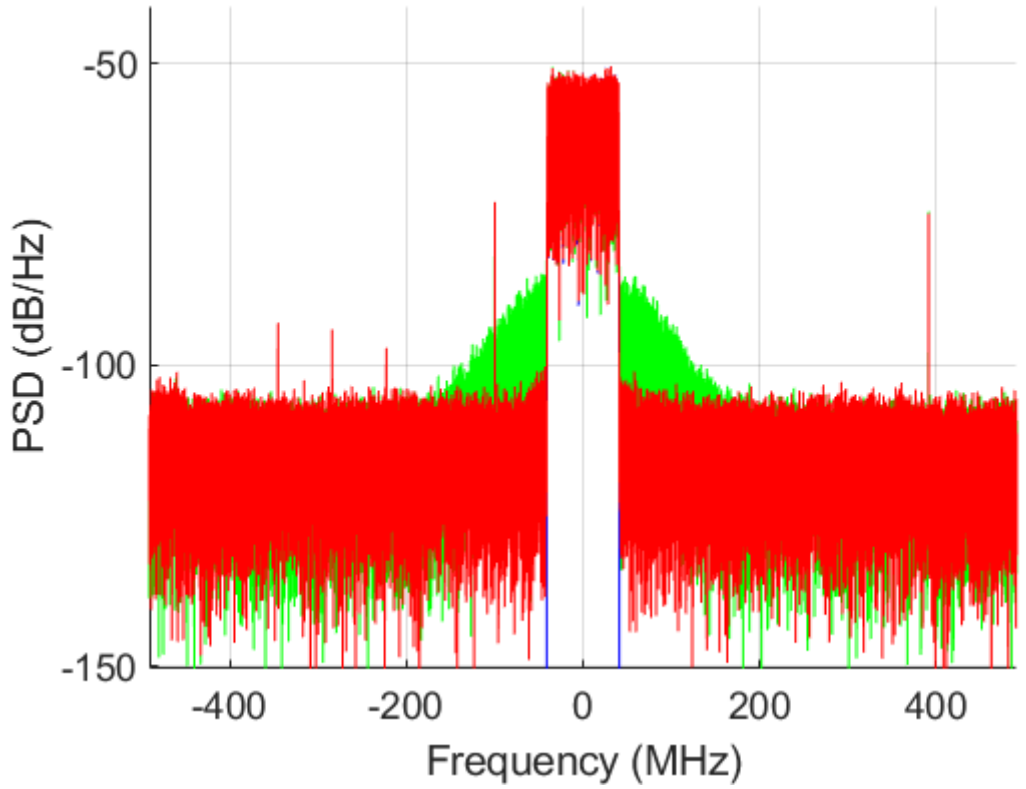

Figure 6.5: Frequency domain representation of experimental technique achieved

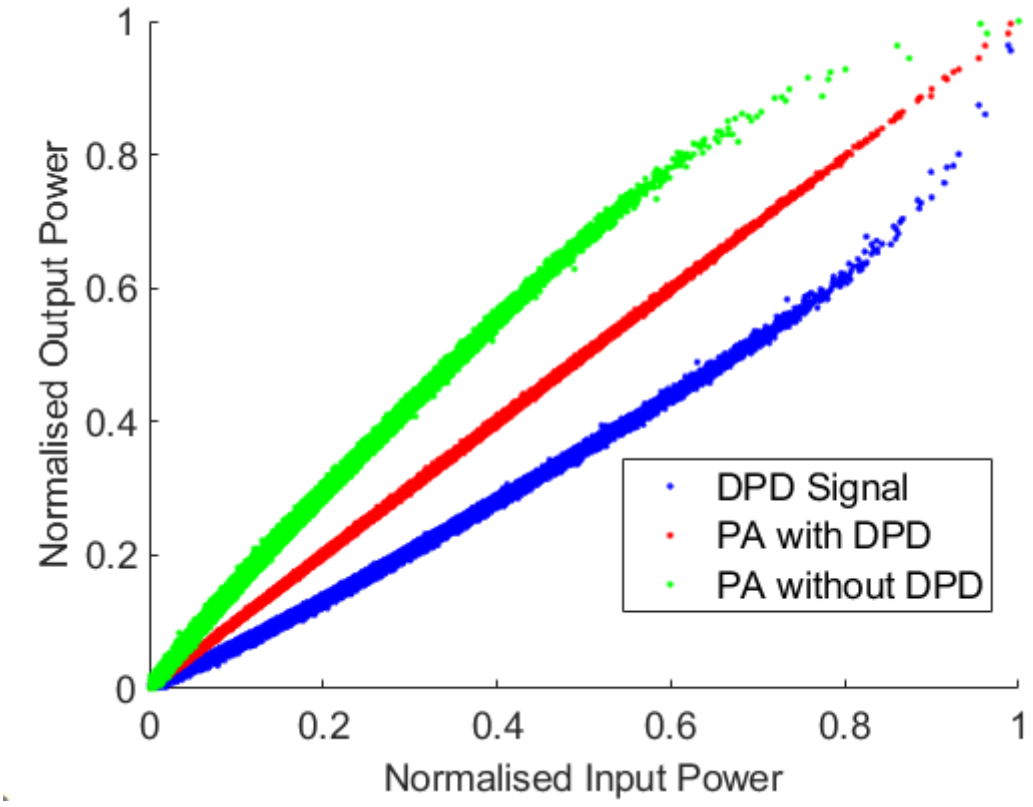

Figure 6.6: 40MHZ 5G NR DPD

Table 6.2: Matrix size versus DPD performance

Training samples	**DS Factor**	**DPD performance (dB)**	**EVM (dB)**
81871	0	17.932168	-46.53416373
40936	2	17.68908529	-46.34859908
20468	4	17.95599863	-46.58461378
10234	8	17.89889682	-46.48277621
5117	16	Failed	
2559	32	Failed	

As can be seen in Figures 6.5 and 6.6, DPD was successful in the instances where the novel metric above, as in equation (6.7) was maintained. Further experimental results can be seen in Table 6.2. Please note DS Factor refers to the down sampling factor in the following Tables. As soon as the metric is no longer consistent with the original signal the DPD coefficient calculation does not work.

As can be seen in Table 6.2 downsampling by a factor of 16 and 32 did not achieve intelligible results. DPD failed in these instances as the variance experienced by the original signal was not maintained in the downsampled signal, as can be seen in Tables 6.3 and 6.4. The coefficient of variation must be maintained.

Table 6.3: Statistical analysis of the real elements of the 5G NR input signal

DS Factor	**Mean**	**Variance**	**Coefficient of Variation**
0	0.003912575	0.031518749	0.022038302768108
2	0.003912575	0.031519133	0.022038168326173
4	0.003912575	0.031519903	0.022037899269155
8	0.003912575	0.031521442	0.022037361233227
16	0.003912574	0.031399881	0.022033142434358
32	0.009240625	0.032141139	0.046249765018970

Table 6.4: Statistical analysis of the imaginary elements of the 5G NR input signal

DS Factor	**Mean**	**Variance**	**Coefficient of Variation**
0	0.001112662	0.03171103	0.006248244681191
2	0.001112662	0.031711417	0.006248206557532
4	0.001112662	0.031712191	0.006248130126848
8	0.001112662	0.03171374	0.006247977560437
16	0.001112662	0.031651637	0.006239973825601
32	0.002448743	0.032485888	0.012241191309849

As can be seen from Tables 6.3 and 6.4 the variance in both the imaginary and real parts of the input signal was not maintained. There is a sudden decrease in the

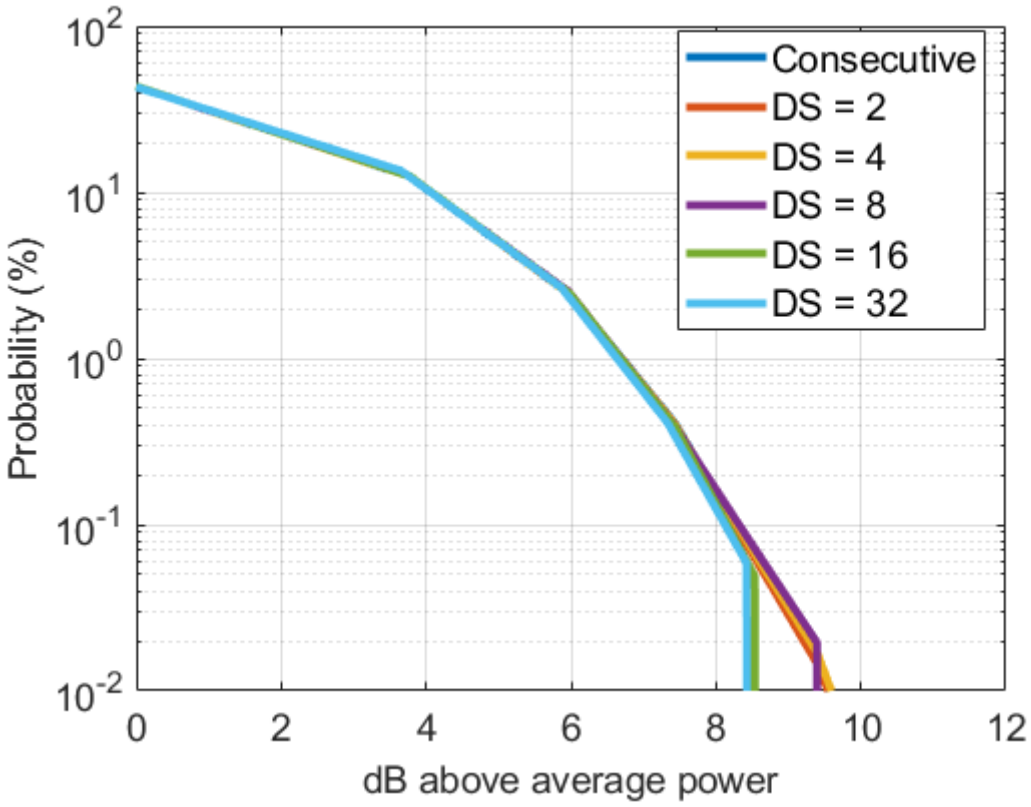

Figure 6.7: PAPR of 5G NR 40MHz input and its downsampled versions

variance of the real part of the signal between the downsampling factors 8 and 16, indicating a loss of peaks.

The coefficient of variation in both parts of the signal (real and imaginary) experience a decrease of $\approx 0.4\%$ and $\approx 0.2\%$ respectively. As can be seen from the Tables 6.3 and 6.4 the mean of the signal is relatively consistent thus the DPD is hindered by the decrease in variance. The result of this decrease of variance ultimately leads to a drop in PAPR as can be seen in Figure 6.7. This sudden decrease in EVM is due to the sensitivity of the previously discussed equation (6.3), without consideration of the number of samples used to train as in equation (6.4), where $\frac{1}{N-1}$ is a scalar value. As this scalar value reduces, the reliance of the mean and variance increases.

Considering the downsampled signals downsampled by eight and sixteen. The matrix size is doubled reducing N from 10233 to 5116 increasing the scaling factor, $\frac{1}{N-1}$, from 9.7723×10^{-5} to 1.9547×10^{-4}. This effect on the standard deviation, or decrease in variance, is more prominent the more limited the number of samples

used.

6.5 Experimental validation : PA behavioural modeling

To validate the performance of the proposed training signal dataset for behavioural modeling, an experimental test bench was developed consisting of the AD9375 transmitting into a 4W Skyworks PA, as can be seen in Figure 6.8. The transmitted signal as shown in figure 6.9, was a 40MHz single carrier FDD 5G-NR signal, modulated by 256QAM.

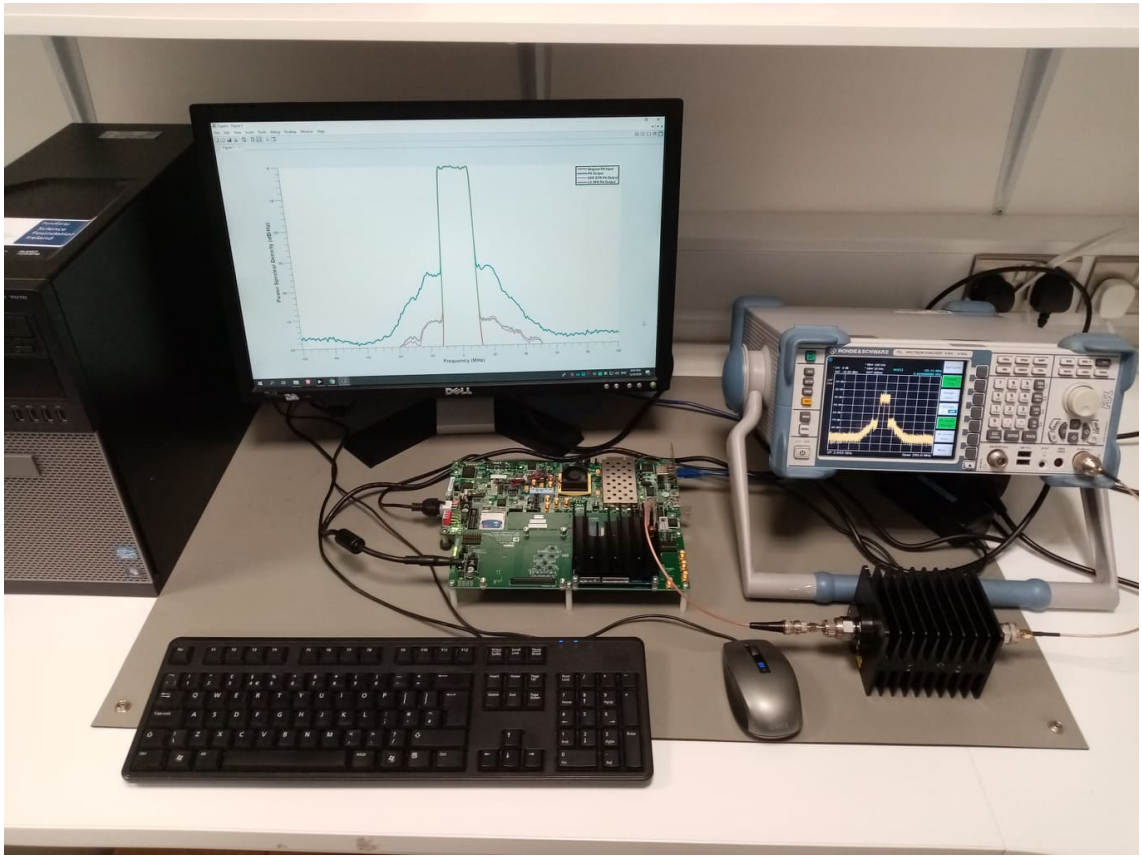

Figure 6.8: Experimental hardware setup for experimentation

Initial experimentation was performed by calculating the highest peak in the large dataset of PA input-output experimental signals, as in Figure 6.9. The MP model was adopted to validate training surrounding the highest peak by consecutively sampling and downsampling by factors of two retaining the same dimension of the

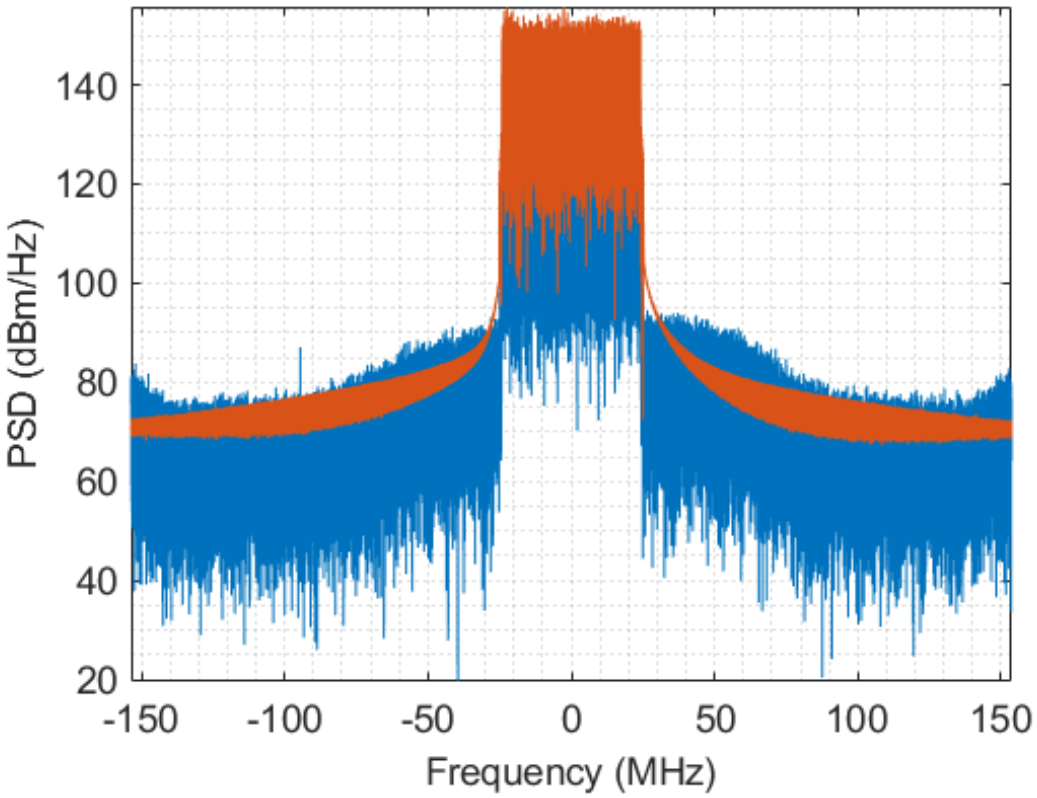

Figure 6.9: 40 MHz 5G-NR 256 QAM single carrier input signal to PA versus output signal of PA, containing 204444 samples, sampled at 307.2 MHz

data matrix.

Figure 6.10 illustrates that consecutive sampling around the highest peak did not result in the most advantageous NMSE, manifesting in the worst NMSE results of the selected training sets. Figure 6.10 also emphasises that downsampling by the highest factor is not necessarily the most beneficial.

Table 6.5: Statistical analysis of 5G 40MHz signal

DS Factor	**Mean**	**Variance** $\times(10^7)$	$\frac{\mu}{\sigma}$	**NMSE (dB)**	$\frac{\Delta\mu}{var}$
0	**15252**	**3.3731**	**2.6261**	**-25.2902**	**1.15**
2	**15044**	**3.4993**	**2.5431**	**-26.6513**	**1.051**
4	**14987**	**3.4795**	**2.5408**	**-27.1767**	**1.0411**
6	**14960**	**3.6143**	**2.4884**	**-26.6805**	**0.9947**
8	**14932**	**3.8301**	**2.4127**	**-25.3911**	**0.93131**
PDF Informed	**14925**	**3.8304**	**2.4115**	**-26.3911**	**0.92937**
Input signal	**11365**	**3.7467**	**1.8567**	**N/A**	**3.0484**

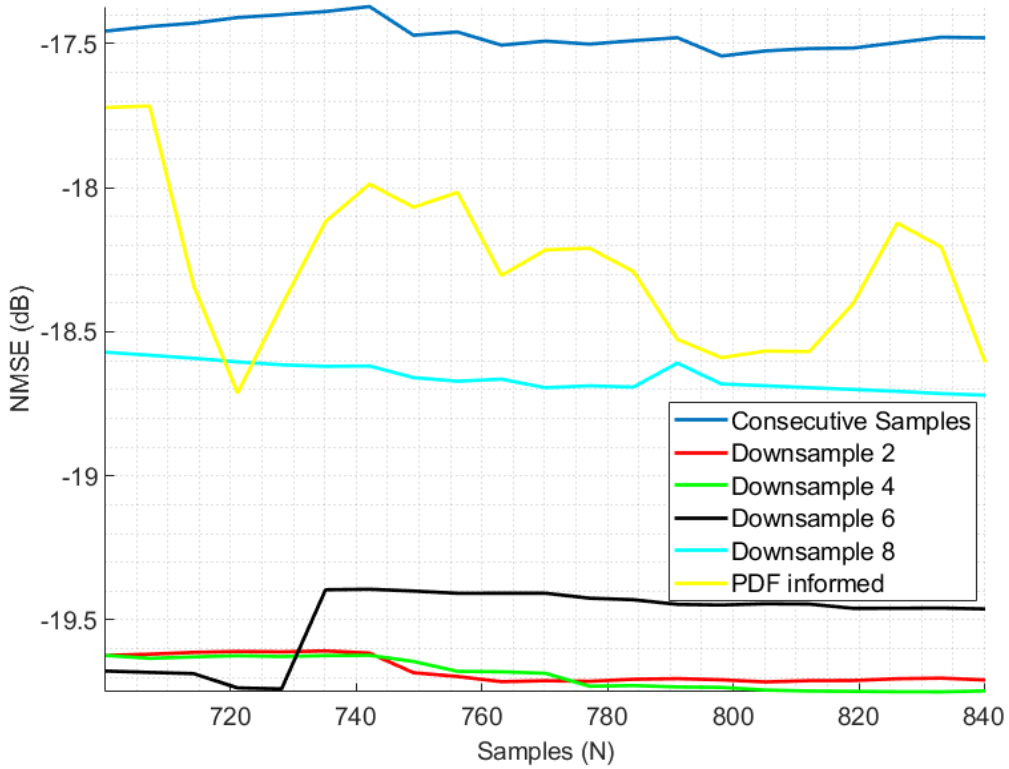

Figure 6.10: NMSE results comparing pre-processing when sampling around the highest peak

Unlike the above DPD experiment, in this case the metric has become a mean biased estimator, such that the above metric given by 6.7 must take into consideration any mean biases. For a mean biased data set the rate of change of the relationship between the mean and variance would ideally increase proportionally, such that $\mu = \sigma^2$. A tandem increase in the variance and mean would entail that although the mean is altering, the variance is altering to account for this change as seen in equation (6.9).

$$\frac{\Delta\mu}{\sigma^2(x)} = 1 \tag{6.9}$$

A comparison was made for Figure 6.11 to illustrate different pre-processing techniques to minimise the number of samples. As can be seen all methods converge over time to a similar NMSE. The PDF informed and downsampling methods perform better over a small data matrix size. As can be seen in Table 6.5, the mean bias

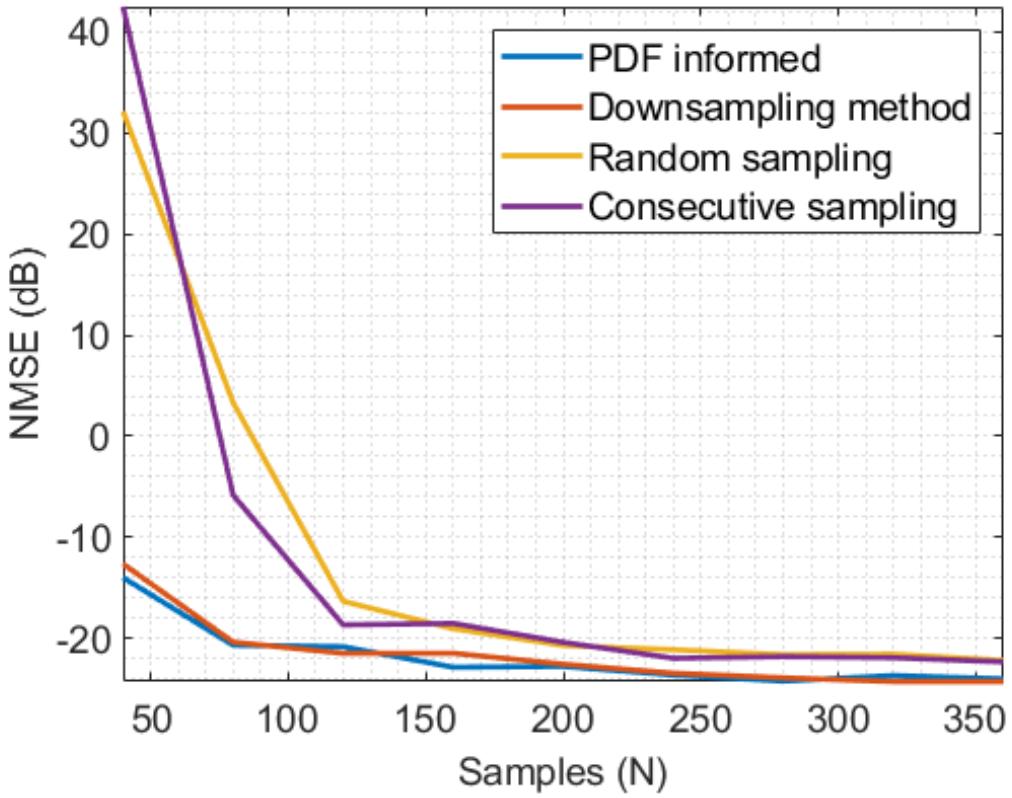

Figure 6.11: NMSE comparison of alternate pre-processing techniques. Please note the downsampling factor used for this data set was 8

must be accounted for to ensure a proportional relationship of μ and the variance experienced by the data set. The NMSE is lowest closest should the metric 6.8 be met.

6.6 Conclusion

In conclusion, this chapter provides a metric to establish the level of pre-processing of an input signal to a DPD/Behavioural model to ensure its output signal retains sufficient similarity to enable accurate modeling of coefficients to be used in DPD or behavioural modeling. This was shown using downsampling as the proposed pre-processing. The metric above was proven for use in pre-processing other than downsampling and experimentally validated on two separate testbenchs.

Further work is planned in order to use the CORDIC algorithm to efficiently im-

plement an algorithm to perform the above metric on an FPGA. CORDIC is a hardware-efficient iterative method that uses rotations to calculate simple functions [109].

Chapter 7

Efficient FPGA Implementation for DPD

Wireless systems such as cellular networks have begun to see proposals for increased operational flexibility through reuse of the same hardware but with different signal standards. Current base stations are static regarding a set carrier frequency and bandwidth capabilities.

This chapter presents an approach to characterise a PA for multiple signal standards, inspired by the concept of Software Defined Radio (SDR). Behavioural modeling is demonstrated to prove that the same coefficients trained for a specific signal standard can be effectively applied to alternate signal standards.

The result of the behavioural modeling is then used to design and implement DPD capable of linearising different signal standards on a Field Programmable Gate Array (FPGA). This implementation is experimentally validated on a state-of-the-art RFSoC FPGA from Xilinx.

In this chapter, AM/AM curves are employed to demonstrate the degree to which different signal standards cause different characteristic behaviour in the power amplifier despite the signals having the same average output power level.

7.1 Related Works

Following on from the discussion in chapter 3.2.1 about Look Up Tables (LUT) a PA behavioral model referred to as Hybrid Look-Up Tables (H-LUT) has been shown to improve the performance relative to a conventional LUT model [110].

FPGA-based LUT solutions have been proposed for switchmode PAs [111]. Authors of [111] demonstrate a polar configured Class F switch-mode PA shown to be effectively linearised. Additional enhancements to the LUT implementation have been proposed [112, 113]. In [112, 113] the authors focus was to reduce the hardware resources required to implement the solution in an FPGA. Molina et al [114] research achieves linearisation with LUTs by expressing the DPD function as a system of linear-in-parameter equations. LS is used to train the LUT coefficients directly. More recent work on LUT [115] is based on spline-interpolated LUTs.

The use of LUT predistorters has been adopted for optical communications. Implementations for nonlinear weighted look-up-table predistortion [116] and reduced size LUT [117] show the continued interest in LUT based predistorters.

Cubic spline basis is a nonlocal basis which has a unique extrapolation properties and is shown to be shown to be a computationally efficient in works by Naraharisetti et al [118]. The research conducted in [118] stores 1-D basis functions in LUTs and calculates 2-D basis functions using tensor product multiplier, enabling a low-cost high-speed calculations of co-efficients not withstanding higher dimensions such as 2-D and 3-D DPD.

Considerable research has been conducted into cognitive radios [119]. SDR's require signals to be transmitted and received linearly, efficiently and flexibly. Traditional DPD/modeling approaches are inefficient in their flexibility with a static carrier frequency and set bandwidths, limiting the signals to be a specific standard.

7.2 Theoretical background

A common technique used to illustrate the characteristic behaviour of a PA is the AM/AM curve. The AM/AM curve depicts the relationship between the input amplitude to the amplifier and the measured output amplitude. The AM/AM curve is commonly utilised to illustrate the nonlinear characteristics of a PA, presenting the distortion in a visually appreciable manner.

Initial research of the methodology proposed in this chapt waser done on the following test bench, as seen in Figure 7.1.

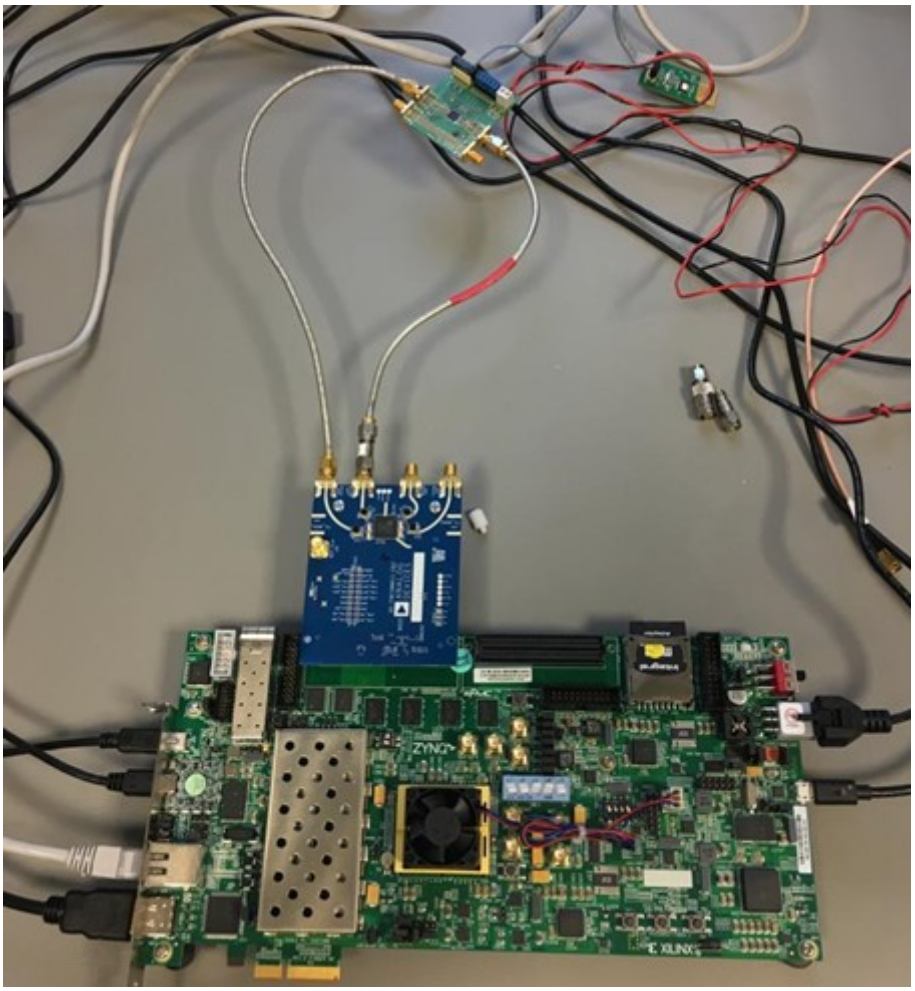

Figure 7.1: PA testbench

The FMCOMMS3 board attached to the ZC706 was used as the RF front end with integrated AD9361 chip from Analog Devices. The PA under test is RFHIC RTP26010-N1 PA. As the output power of the FMCOMMS3 is limited, a driver PA

BGA7210 is used to increase the input power to the test PA. The BGA7210 is a high linearity PA with variable gain and it is operating at linear region.

Investigations into PA characterisation given different input signals led to the observation that various AM/AM curves of various standards, at different power levels, are very similar.

7.3 Experimental validation

Multiple signals were required for each of the standards studied. The signals used were generated in MATLAB using modulation functions from the Communications, LTE & 5G toolboxes. Measured results were obtained using the following testbenchs as seen in Figures 7.2 and 7.6. Here an RFSoC ZCU111 for transmitting and receiving the signals has been used. The signal was generated at the Intermediate Frequency (IF) centred at 1GHz from the RFSoC DAC with sampling frequency of 737.28 MHz and then with the upconverted to the required center frequency of 2.6 GHz. The signal was then transmitted through a 10W GaN-SiC pallet amplifier (RFHIC RTP26010-N1) and the power of the signal was maintained sufficiently to drive the PA in a nonlinear region of operation. This PA has two output ports, one of which was connected to a spectrum analyser and the second coupled output port was connected to the downconverter mixer which uses the same LO frequency as of the upconverter mixer. The signal is then downconverted back to the IF frequency i.e., 1GHz and is passed to the RFSoC ADC at a sampling rate of 737.28 MSPS. In the case of this PA the model memory depth was chosen to be 3 and the non-linear order of the MP model was chosen to be 3.

For measuring at different power levels, the power of the signals generated from the RFSoC were adjusted using the RF Data Converter Interface. The sample length generated in MATLAB for standardised signals was approximately 70,000 samples. To time align the signal and reduce the noise floor to achieve better dynamic range, the length of the signal captured was 10 times the length of the transmitted signal

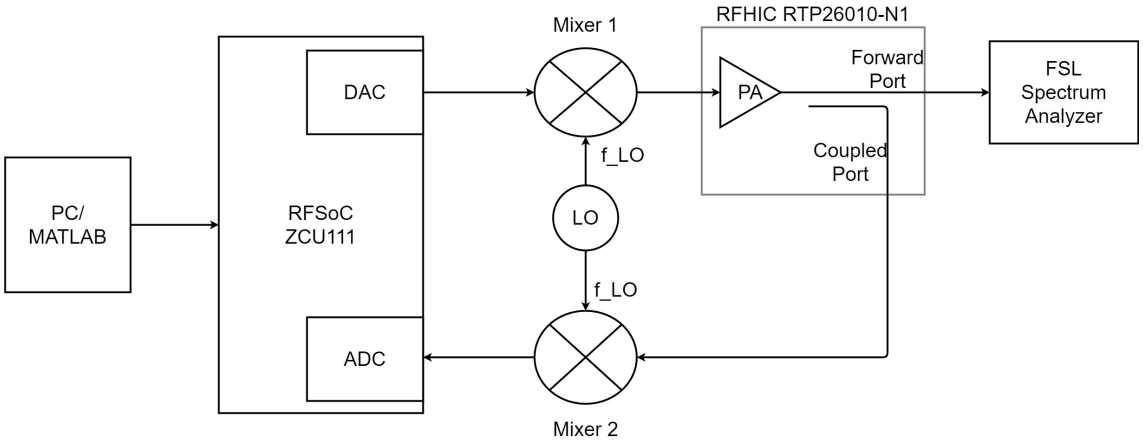

Figure 7.2: Test bench block diagram

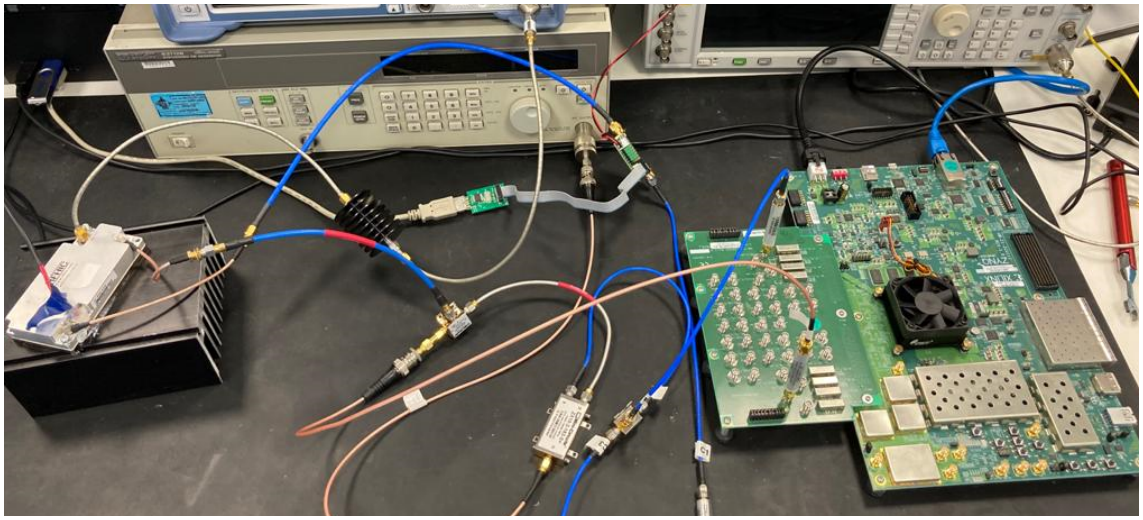

Figure 7.3: Experimental Measurement Bench RFSoC ZCU111 with RFHIC RTP26010-N1 PA

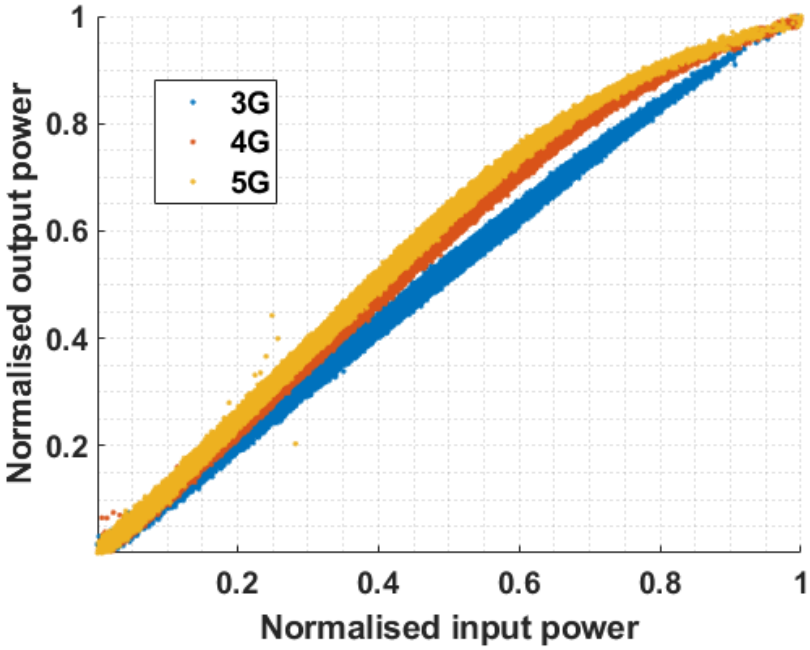

Figure 7.4: An experimentally validated illustration of the input-output signal relationship transmitted at 2.6GHz for 3G,4G and 5G signals. Signals were transmitted through the same PA at equivalent transmit power.

i.e., 700,000 samples. Once the signal has been time aligned, both the transmitted and received signals were normalised with respect to the maximum absolute value of each. The first 30,000 samples of the averaged input and output signals have been used for training the behavioural model and a further 30,000 were used for the validation of the model.

In this work, AM/AM curves are employed to demonstrate the degree to which different signal standards cause different characteristic behaviour from the power amplifier when signals have the same output power level. Figure 7.4 shows the discrepancies between experimentally measured 3G, 4G and 5G signals with the same power level, passed through a PA operating in the nonlinear region.

Further investigations into PA characterisation given different input signals led to the observation that various AM/AM curves of 3G, 4G and 5G-NR, at different combinations of power levels, are similar, as seen in Figure 7.5. Therefore by noting the relative difference in the signal power levels for the different signal standards, that all yield similar characteristic AM/AM performance, model coefficients are

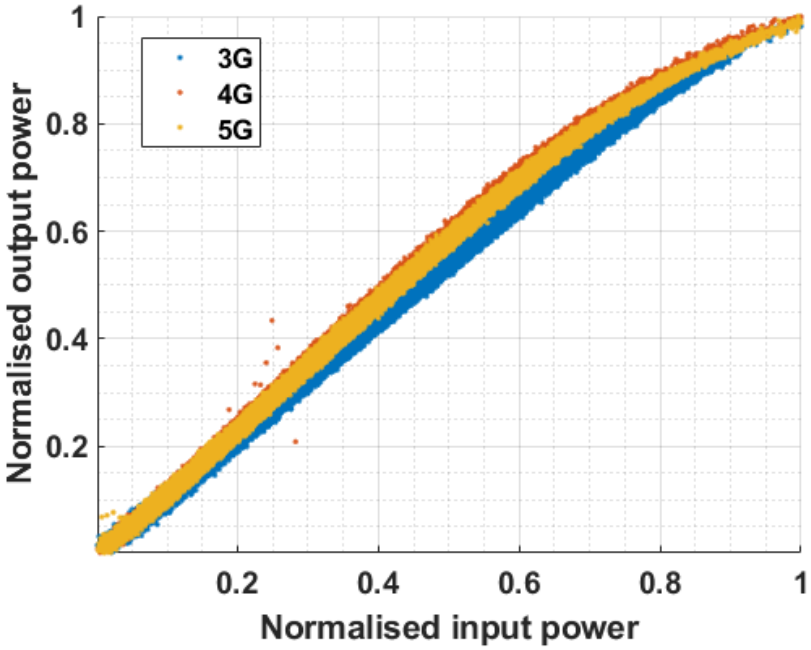

Figure 7.5: Experimentally validated AM/AM curve illustrating that 3G, 4G and 5G input signals can produce comparable AM/AM curves given the signals are sent at disparate transmit power levels.

stored.

As indicated in Figure 7.4, for the same power amplifier excited by different signals with the same average power level and similar bandwidth, there is a noticeable difference in the AM/AM curves. This in turn indicates a difference of behaviour of the hardware as a result of the different signals.

It is therefore not sufficient to assume a particular operating behaviour for the PA based on the average output signal power alone; the input signal transmitted must also be taken into consideration.

This chapter presents a means by which a single set of behavioural model coefficients can be extracted and stored in a look-up-table for use with any signal standard so long as the relative power level offsets between different signal types are accounted for. The matching of behavioural model coefficients is implemented across different signal standards. This is performed by extracting coefficients for different power levels for each standard. From the measured output signals the corresponding AM/AM curves are fitted. One set of coefficients are extracted for one signal standard and

the AM/AM curve for that standard which best fits the other signal standards is sought.

By exciting the power amplifier using various signals at a range of operating power levels, sets of signals can be compiled and relationships between standards can be learned. Importantly, the signals which match closest in terms of model coefficient performance do not have identical operating power levels.

In order to validate the proposed technique for multiple signal standards, commonly used signal modulation schemes for 3G, 4G and 5G communications are examined. A signal bandwidth of 20 MHz for single carrier signals were sent through an RFHIC Doherty PA as described above.

Table 7.1: Relationship Between Signal Standards and Power Levels

Coefficient sets	1	2	3	4	5	6	7	8
5G (dBFS)	-10	-9	-8	-7	-6	-5	-4	-3
4G (dBFS)	-18	-17	-16	-15	-14	-13	-12	-11
3G (dBFS)	-20	-19	-18	-17	-16	-15	-14	-13

Table 7.1, illustrates 8 sets of power levels in dBFS that are most similar, for three signal standards.

Pairs of input and output signals were captured for both PAs. The AM/AM curves for one signal standard were plotted as a visual aid. A subset of samples from the alternative signal standards at similar power levels were subsequently used to check if their AM/AM trace follows a similar trajectory. Comparing these, the relative power levels between standards is determined, and the model coefficients are indexed in the LUT relative to each signal standard power level. The relative power level offsets between standards can be seen in Table 7.1. In this instance, eight sets of

Table 7.2: Cross Signal Standard Model Accuracy NMSE (dB)

$\frac{Training\rightarrow}{Test\downarrow}$	3G MP/GMP	4G MP/GMP	5G MP/GMP
3G	-38.386 / -38.39	-42.008 / -41.931	-41.856/-41.804
4G	-42.02/-41.961	-45.874 / -45.879	-41.995 / -41.918
5G	-42.178 / -42.13	-39.623/ -30.589	-45.874/-44.067

coefficients are matched for the same power amplifier across three different signal standards which each have a relative power level offset.

Using the relative power level offsets, a behavioral model can be trained for one signal standard and reliably used to model the PA response across the other signal standards. Table 7.2 illustrates the NMSE comparisons calculated between different signal standards with similar AM/AM curves, which corresponds to one of the columns in Table 7.1. The columns of Table 7.2 are populated by training the coefficients using one of the signal standards and validating the model accuracy for all three signal standards. The accuracy for each standard is given in NMSE (dB) and placed in its respective row. Independent output signals which were not used to train the models were used for the validation in each case.

7.4 DPD implementation

The DPD was performed on an RFSoC ZCU111. The signal was generated at the Intermediate Frequency (IF) centred at 2.6GHz with a sampling frequency of 737.28 MHz. The signal was passed through the GaN-SiC pallet amplifier (RFHIC RTP26010-N1) and the power of the signal was maintained sufficiently to drive the PA in a nonlinear region of operation. This particular PA has two output ports, one of which was connected to a spectrum analyser and the second coupled output port was connected to the RFSoC ADC which also has a sampling rate of 737.28 MSPS. In the case of this PA the model memory depth was chosen to be 3 and the non-linear order of the MP model was chosen to be 3.

Three MP DPD models are obtained and each model is trained with a particular signal standard. To validate performance of the resulting DPD coefficients, each set of extracted coefficients are tested for three signal standards, namely 3G, 4G and 5G. From these experiments on the PA hardware, Table 7.3 shows the NMSE performance for different signal standards. In each row, the test signal is the same, independent of the training model it is applied to. This includes the power level.

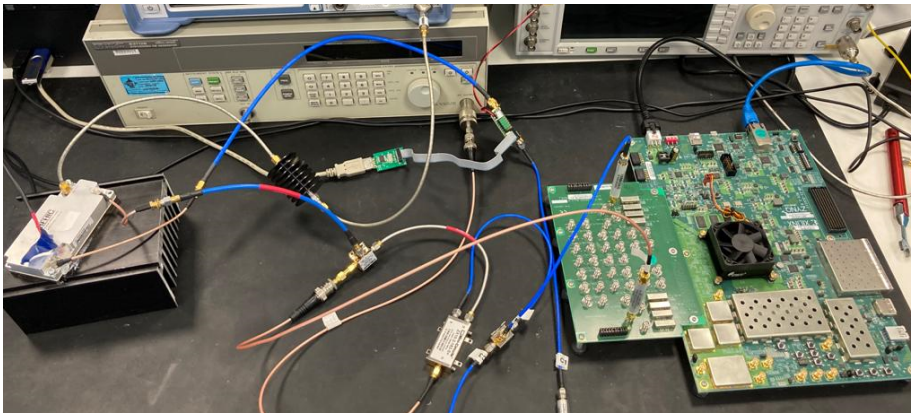

Figure 7.6: Experimental Measurement Bench RFSoC ZCU111 with RFHIC RTP26010-N1 PA

Table 7.3: NMSE Comparison of proposed DPD and conventional DPD

Power level(dBm)	6	6.5	7	7.5	8	8.5
proposed DPD NMSE (dB)	-25.22	-27.59	-28.30	-28.64	-28.91	-30.76
conventional DPD NMSE (dB)	-28.57	-28.73	-29.86	-29.55	-30.84	-31.94
difference NMSE (dB)	-3.35	-1.14	-1.56	-0.91	-1.93	-1.18

With the same test signal we achieve similar performance across different standards. The results show that the DPD model can maintain similar performance even when the test signals and training signal are under different standards.

A second testbench, designed by colleagues in North Eastern University, was used to validate the proposed methodology and can be seen in Figure 7.7.

The testbench setup shown in Fig. 8 using Xilinx's RFSoC Gen 3 (ZCU216) is used to perform DPD and baseband signal processing. Using both software and hardware design, the address selection block is implemented on the ARM core which utilizes the AXI-Lite bus to update the coefficients based on the input signal. The computationally intensive task of computing the memory polynomial model is performed on FPGA fabric. The test PA is an AFSC5G37D37 Doherty PA from NXP.

7.5 Chapter Summary

The work presented in this chapter demonstrates what set of coefficients to use based on the observed nonlinear performance of the PA, for any given signal. This can

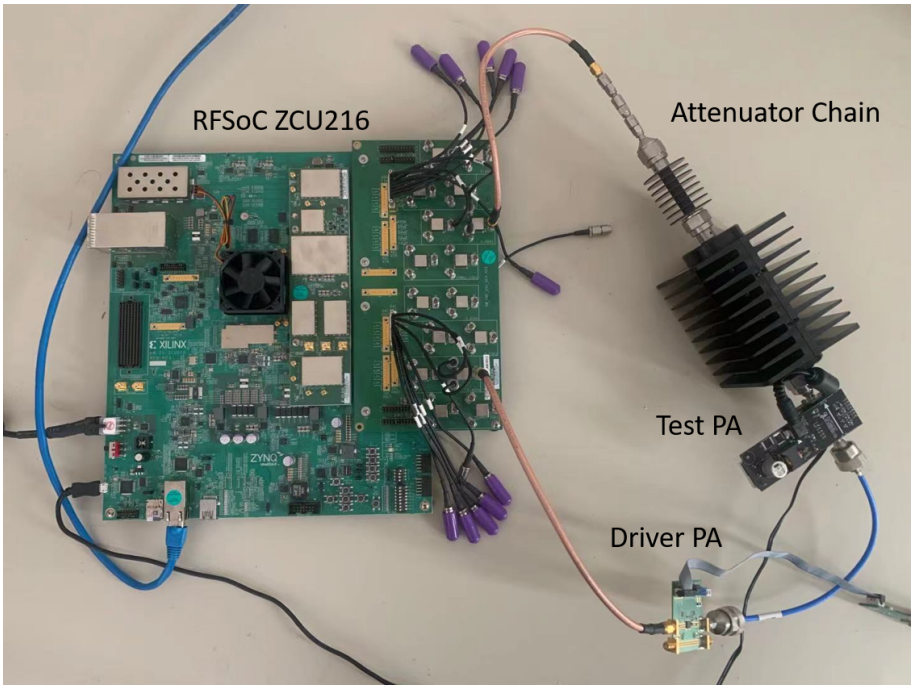

Figure 7.7: DPD TestBench RFSoC ZCU216 with NXP's AFSC5G37D37 Doherty PA

avoid the need for the standard LUT approach which has to train a set of coefficients for every possible combination of signal standard and operating power level.

This chapter provides a definitive solution to behavioural modeling and digital pre-distortion for multiple signal standards using LUTs. By matching the relative AM/AM curves for different signals passed through the same power amplifier, sets of common coefficients that will work across different signal standards can be found.

While the technique is demonstrated using polynomial models and Least Squares, the relationships between different input signal standards exist irrespective of the model structure used.

Experimental validation is performed using input signals of three different signal modulation schemes, and behavioral modelling and DPD are carried out with experimental measurements using two different PAs from two different manufacturers. The results show that training of a model, the most computationally intensive aspect, can be done for one signal standard and successfully applied to others provided

the relative signal power level offsets are known.

Chapter 8

Conclusion

The contribution of this thesis is to improve polynomial behavioural modeling and DPD techniques with a particular focus on improving training and dimension reduction techniques.

Contributions in Chapter 4 address the implementation of an early stopping technique, to avoid instability, of the Recursive Least Squares technique when training memory polynomial based behavioural models. Experimental validation of the proposed procedure shows that the NMSE of the experimental output vs estimated output indicates high fidelity until the point identified by the early stopping criterion is exceeded after which point it deteriorates rapidly.

The contributions in Chapter 5 develop a novel DPD function for the purposes of increasing computational efficiency of updating coefficients locally, enabling matrix inversions to be more compact and achieve high performance. Chapter 5 demonstrates this method for selective partitioned adaptation for a LS trained DPD system. The novel methodology presented exploited the Frisch-Waugh-Lovell technique to enable a more accurate adaptation. The partitioned regression was accomplished by applying the FWL theorem to a DPD function trained using LS regression. The proposed methodology was experimentally validated by adapting a particular subset of a GMP DPD function which linearised a PA amplifying a 40MHz 5G-NR signal.

Contributions presented in Chapter 6 support ongoing research to provide a metric to establish the level of pre-processing of an input signal to a DPD/Behavioural model to ensure its output signal retains sufficient similarity to enable accurate modeling of coefficients to be used in DPD or behavioural modeling. This was shown using downsampling as the proposed pre-processing. The metric was proven for use in pre-processing other than downsampling and experimentally validated on two separate testbenchs.

The contribution contained within Chapter 7 demonstrates a novel look-up table indexing technique to help determine the optimal DPD coefficients needed for pre-distortion and behavioural modeling. Experimental validation is performed using input signals of three different signal modulation schemes, and behavioral modelling and DPD are carried out with experimental measurements using two different PAs from two different manufacturers.

8.1 Further Work

Further research is planned to investigate various research questions arising from this thesis.

- Further work will be targeted in Chapter 5 to investigate more fully how the FWL partition could be optimally determined. My point of view is that this could be very useful in use cases such as long term memory effects could be mitigated against by partitioned coefficient over alternate iterations of DPD.
- Concerning the metric used to determine the validity of pre-processing in Chapter 6, it is planned to implement an algorithm on an FPGA to minimise the number of data samples while maintaining high fidelity using the CORDIC algorithm. Alternatively research could be undertaken to use the same statistical analysis in order to enhance the speed of DPD , and the maximum standard deviation of a signal pre-transmission is known and can be assessed using it's modulation technique.

Bibliography

[1] M. W. Akhtar, S. A. Hassan, R. Ghaffar, H. Jung, S. Garg, and M. S. Hossain, "The shift to 6G communications: vision and requirements," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, 2020. [Online]. Available: https://doi.org/10.1186/s13673-020-00258-2

[2] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5g," *IEEE Network*, vol. 30, no. 1, pp. 44–51, January 2016.

[3] I.T.Union, "Imt vision - framework and overall objectives of the future development of imt for 2020 and beyond," *International Telecommunications Union*, 2015.

[4] K. Santhi, V. Srivastava, G. SenthilKumaran, and A. Butare, "Goals of true broad band's wireless next wave (4g-5g)," in *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No.03CH37484)*, vol. 4, 2003, pp. 2317–2321 Vol.4.

[5] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[6] A. S. Tehrani, T. Eriksson, and C. Fager, "Modeling of long term memory effects in rf power amplifiers with dynamic parameters," in *2012 IEEE/MTT-S International Microwave Symposium Digest*, 2012, pp. 1–3.

[7] M. Steer, *Microwave and RF design : a systems approach*. SciTech, 2013.

[8] L. B. Milstein, "Wideband code division multiple access," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 8, pp. 1344–1354, Aug 2000.

[9] A. R. Mishra, *Fundamentals of network planning and optimisation 2G/3G/4G: evolution to 5G*, second edition ed. Hoboken, NJ, USA: Wiley, 2018.

[10] T. Akhtar, C. Tselios, and I. Politis, *Radio resource management: approaches and implementations from 4G to 5G and beyond*. Springer US, 2021, vol. 27, no. 1. [Online]. Available: https://doi.org/10.1007/s11276-020-02479-w

[11] M. Akurati, Y. Kamatham, S. K. Pentamsetty, and S. Prasad Kodati, "Papr reduction in ofdm using hybrid companding for 5g wireless communications," in *2019 Global Conference for Advancement in Technology (GCAT)*, 2019, pp. 1–5.

[12] A.-F. B. Ali Bachir, M. Zhour, and M. Ahmed, "Modeling and design of a dvb-s2x system," in *2019 5th International Conference on Optimization and Applications (ICOA)*, 2019, pp. 1–5.

[13] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Pothecary, J. F. Sevic, and N. O. Sokal, "Power Amplifiers and Transmitters for RF and Microwave," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 814–826, 2002.

[14] P. Roblin, S. J. Doo, X. Cui, G. H. Jessen, D. Chaillot, and J. Strahler, "New ultra-fast real-time active load-pull measurements for high speed rf power amplifier design," in *2007 IEEE/MTT-S International Microwave Symposium*, 2007, pp. 1493–1496.

[15] P. B. Kenington, *High Linearity RF Amplifier Design*, 1st ed. USA: Artech House, Inc., 2000.

[16] T. Johansson and J. Fritzin, "A Review of Watt-Level CMOS RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 1, pp. 111–124, 2014.

[17] Y. Liu, J. J. Yan, H. Dabag, and P. M. Asbeck, "Novel Technique for Wideband Digital Predistortion of Power Amplifiers With an Under-Sampling ADC," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2604–2617, 2014.

[18] P. Roblin, C. Liang, H.-C. Chang, and K. Rawat, "Class-e pa prototype using an embedding model: Invited paper," in *2019 IEEE 20th Wireless and Microwave Technology Conference (WAMICON)*, 2019, pp. 1–6.

[19] S. Madhuwantha, P. Ramabadran, R. Farrell, and J. Dooley, "N-way digitally driven doherty power amplifier design and analysis for ku band applications," in *2018 29th Irish Signals and Systems Conference (ISSC)*, 2018, pp. 1–6.

[20] M. Mortazavi, Y. Shen, D. Mul, L. C. N. de Vreede, M. Spirito, and M. Babaie, "A four-way series doherty digital polar transmitter at mm-wave frequencies," *IEEE Journal of Solid-State Circuits*, pp. 1–1, 2022.

[21] G. T. Watkins, "The Best of Both Worlds: The Dynamic Load-Modulation Power Amplifier," *IEEE Microwave Magazine*, vol. 21, no. 4, pp. 76–86, 2020.

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

[23] R. McMorrow, D. Upton, and P. Maloney, "The microwave doherty amplifier," in *1994 IEEE MTT-S International Microwave Symposium Digest (Cat. No.94CH3389-4)*, 1994, pp. 1653–1656 vol.3.

[24] J.-y. Lee, J.-y. Kim, J.-h. Kim, K.-j. Cho, and S. P. Stapleton, "A high power asymmetric doherty amplifier with improved linear dynamic range," in *2006 IEEE MTT-S International Microwave Symposium Digest*, 2006, pp. 1348–1351.

[25] C. Jebali and A. Kouki, "Thermal effects analysis of gan hemt power amplifier based on ltcc substrate integration," in *2018 IEEE Canadian Conference on*

Electrical Computer Engineering (CCECE), 2018, pp. 1–4.

[26] M. T. Ozalas, "The impact of electro-thermal coupling on hbt power amplifiers," in *2014 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, 2014, pp. 1–4.

[27] J. C. Pedro and N. M. G. B. de Carvalho, "Intermodulation distortion in microwave and wireless circuits," 2003.

[28] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of twt amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, November 1981.

[29] O. Hammi, F. M. Ghannouchi, S. Boumaiza, and B. Vassilakis, "A data-based nested LUT model for RF power amplifiers exhibiting memory effects," *IEEE Microwave and Wireless Components Letters*, vol. 17, no. 10, pp. 712–714, oct 2007.

[30] O. Hammi, F. M. Ghannouchi, and B. Vassilakis, "2-D vector quantized behavioral model for wireless transmitters' nonlinearity and memory effects modeling," in *2008 IEEE Radio and Wireless Symposium, RWS*, 2008, pp. 763–766.

[31] L. C. Nunes, P. M. Cabral, and J. C. Pedro, "LUT Based Behavioral Model for Doherty Power Amplifier Design," in *Conf. on Telecommunications - ConfTele*, sep 2015, pp. –.

[32] M. Hoflehner and A. Springer, "Comparison of rf power amplifier behavioral models with respect to their modeling capabilities in adjacent and alternate bands," in *Computer Aided Systems Theory – EUROCAST 2011*, R. Moreno-Díaz, F. Pichler, and A. Quesada-Arencibia, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–16.

[33] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 4, pp. 1150–1163, 2005.

[34] A. Zhu, J. Dooley, and T. J. Brazil, "Simplified volterra series based behavioral modeling of rf power amplifiers using deviation-reduction," in *2006 IEEE MTT-S International Microwave Symposium Digest*, June 2006, pp. 1113–1116.

[35] B. Fehri and S. Boumaiza, "Baseband equivalent volterra series for digital predistortion of dual-band power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 3, pp. 700–714, 2014.

[36] S. Benedetto, E. Biglieri, and R. Daffara, "Modeling and Performance Evaluation of Nonlinear Satellite Links-A Volterra Series Approach," *IEEE Transactions on Aerospace and Electronic Systems*.

[37] A. Hekkala, M. Hiivala, M. Lasanen, J. Perttu, L. C. Vieira, N. J. Gomes, and A. Nkansah, "Predistortion of Radio Over Fiber Links: Algorithms, Implementation, and Measurements," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 3, pp. 664–672, 2012.

[38] N. Lashkarian and C. Dick, "FPGA Implementation of Digital Predistortion Linearizers for Wideband Power Amplifiers," *Proc. of SDR*, 2004.

[39] J. Harmon and S. G. Wilson, "Iterative Approach to the Indirect Learning Architecture for Baseband Digital Predistortion," in *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, 2010, pp. 1–5.

[40] H. Ku and J. S. Kenney, "Behavioral Modeling of RF Power Amplifiers Considering IMD and Spectral Regrowth Asymmetries," *Microwave Symposium Digest, 2003 IEEE MTT-S International*, 2003.

[41] N. Messaoudi, M. Fares, S. Boumaiza, and J. Wood, "Complexity Reduced Odd-Order Memory Polynomial Pre-Distorter for 400-Watt Multi-Carrier Doherty Amplifier Linearization," in *2008 IEEE MTT-S International Microwave Symposium Digest*, 2008, pp. 419–422.

[42] L. Ding and G. T. Zhou, "Effects of Even-Order Nonlinear Terms on Power Amplifier Modeling and Predistortion Linearization," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 156–162, 2004.

[43] H. Enzinger, K. Freiberger, and C. Vogel, "Analysis of Even-Order Terms in Memoryless and Quasi-Memoryless Polynomial Baseband Models," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1714–1717.

[44] A. Zhu, J. Dooley, and T. J. Brazil, "Simplified Volterra Series Based Behavioral Modeling of RF Power Amplifiers Using Deviation-Reduction," in *2006 IEEE MTT-S International Microwave Symposium Digest*, 2006, pp. 1113–1116.

[45] C. Crespo-Cadenas, J. Reina-Tosina, M. J. Madero-Ayora, and J. Munoz-Cruzado, "A New Approach to Pruning Volterra Models for Power Amplifiers," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2113–2120, 2010.

[46] V. J. Mathews and G. L. Sicuranza, *Polynomial signal processing*. Wiley, 2000.

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

[48] Y.-J. Liu, J. Zhou, W. Chen, and B.-H. Zhou, "A robust augmented complexity-reduced generalized memory polynomial for wideband rf power amplifiers," *IEEE Transactions on Industrial Electronics*, vol. 61, pp. 2389 – 2401, 10 2013.

[49] F. M. Ghannouchi, O. Hammi, and M. Helaoui, *Behavioral modeling and predistortion of wideband wireless transmitters*. John Wiley Sons, 2015.

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

[51] R. Kline, "Harold black and the negative-feedback amplifier," *IEEE Control Systems Magazine*, vol. 13, no. 4, pp. 82–85, 1993.

[52] R. N. Braithwaite and A. Khanifar, "High Efficiency Feedforward Power Amplifier Using a Nonlinear Error Amplifier and Offset Alignment Control," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013, pp. 1–4.

[53] R. N. Braithwaite, "A Comparison for a Doherty Power Amplifier Linearized Using Digital Predistortion and Feedforward Compensation," in *2015 IEEE MTT-S International Microwave Symposium*, 2015, pp. 1–4.

[54] D. Psaltis, A. Sideris, and A. A. Yamamura, "A multilayered neural network controller," *IEEE Control Systems Magazine*, vol. 8, no. 2, pp. 17–21, April 1988.

[55] Changsoo Eun and E. J. Powers, "A new volterra predistorter based on the indirect learning architecture," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 223–227, Jan 1997.

[56] C. Eun and E. Powers, "A new volterra predistorter based on the indirect learning architecture," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 223–227, 1997.

[57] S. Amin, E. Zenteno, P. N. Landin, D. Rönnow, M. Isaksson, and P. Händel, "Noise impact on the identification of digital predistorter parameters in the indirect learning architecture," in *2012 Swedish Communication Technologies Workshop (Swe-CTW)*, Oct 2012, pp. 36–39.

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

[59] O. Hammi and F. Ghannouchi, "Power alignment of digital predistorters for power amplifiers linearity optimization," *Broadcasting, IEEE Transactions on*, vol. 55, pp. 109 – 114, 04 2009.

[60] K. J. Muhonen, M. Kavehrad, and R. Krishnamoorthy, "Look-up table techniques for adaptive digital predistortion: a development and comparison," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1995–2002, Sep. 2000.

[61] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 120–133, Jan 2007.

[62] X. Liu, W. Chen, L. Chen, and Z. Feng, "A robust and broadband digital predistortion utilizing negative feedback iteration," in *2018 IEEE MTT-S International Wireless Symposium (IWS)*, May 2018, pp. 1–4.

[63] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *2008 IEEE International Symposium on Wireless Communication Systems*, Oct 2008, pp. 309–313.

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

[65] H. Le Duc, B. Feuvrie, M. Pastore, and Y. Wang, "An adaptive cascaded ila- and dla-based digital predistorter for linearizing an rf power amplifier," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 3, pp. 1031–1041, March 2019.

[66] S. Rahmanian, M. Bateni, and E. Yazdian, "Efficient fpga implementation of a digital predistorter for power amplifier linearization," *Circuits, Systems, and Signal Processing*, vol. 39, 04 2020.

[67] J. Ren, "A new digital predistortion algorithms scheme of feedback fir cross-term memory polynomial model for short-wave power amplifier," *IEEE Access*, vol. PP, pp. 1–1, 02 2020.

[68] G. Montoro, P. L. Gilabert, E. Bertran, A. Cesari, and J. A. García, "An LMS-based adaptive predistorter for cancelling nonlinear memory effects in RF power amplifiers," *Asia-Pacific Microwave Conference Proceedings, APMC*, no. May 2014, 2007.

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

[70] S. Puntanen and G. P. H. Styan, "The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator," *The American Statistician*, vol. 43, no. 3, pp. 153–161, sep 1989. [Online]. Available: http://www.jstor.org/stable/2685062

[71] A. Albert, "The Gauss-Markov Theorem for Regression Models with Possibly Singular Covariances," *SIAM Journal on Applied Mathematics*, vol. 24, no. 2, pp. 182–187, sep 1973. [Online]. Available: http://www.jstor.org/stable/2099674

[72] F. Gregorio, J. Cousseau, S. Werner, and T. Riihonen, "Evm analysis for broadband ofdm direct-conversion transmitters," *Vehicular Technology, IEEE Transactions on*, vol. 62, pp. 3443–3451, 09 2013.

[73] K. Finnerty, "Linear operation of switch-mode outphasing power amplifiers," Ph.D. dissertation, 2016.

[74] *RF Power Amplifier Behavioral Modeling*, ser. The Cambridge RF and Microwave Engineering Series. Cambridge University Press, 2008.

[75] R. N. Braithwaite, "Fixed point considerations for digital predistortion of a rf power amplifier using recursive least square (rls) estimation," in *2019*

IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR), 2019, pp. 1–3.

[76] S. Moghaddamnia, M. Fuhrwerk, and J. Peissig, "Use of statistical signal properties for adaptive predistortion of high power amplifiers," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–6.

[77] Z. Jiang, J. Wang, and W. Wu, "A piecewise rls-based digital predistortion scheme for power amplifier," in *Proceedings of the 2nd International Conference on Telecommunications and Communication Engineering*, ser. ICTCE 2018. New York, NY, USA: Association for Computing Machinery, 2018, p. 167–170. [Online]. Available: https://doi.org/10.1145/3291842.3291848

[78] W. Jian, C. Yu, J. Wang, J. Yu, and L. Wang, "OFDM adaptive digital predistortion method combines RLS and LMS algorithm," *2009 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009*, pp. 3900–3903, 2009.

[79] O. Z. Alngar, W. S. El-Deeb, and E.-S. M. El-Rabaie, "Improving the performance of the digital predistorter based on sample reuse- rls algorithm," in *2018 International Japan-Africa Conference on Electronics, Communications and Computations (JAC-ECC)*, 2018, pp. 47–50.

[80] J. N. Swaminathan and P. Kumar, "A Novel ML-2D-LUT Based Adaptive Predistorter of High Power Amplifier Using New Improved RLS Algorithm," *Wireless Personal Communications*, vol. 90, no. 2, pp. 807–816, 2016.

[81] M. Loughman, R. Farrell, and J. Dooley, "Early stopping criteria for adaptive training of dynamic nonlinear behavioural models," in *2019 30th Irish Signals and Systems Conference (ISSC)*. IEEE, 2019, pp. 1–5.

[82] P. S. R. Diniz, *Adaptive filtering*. Springer, 2006.

[83] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications, 2nd Edition.* John Wiley Sons, 2013.

[84] D. C. Lay, S. R. Lay, and J. McDonald, *Linear algebra and its applications.* Pearson, 2015.

[85] X. Lin, Y. Zhang, H. Li, G. Li, W. Qiao, and F. Liu, "Low computational complexity digital predistortion based on independent parameters estimation," in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, 2019, pp. 1501–1505.

[86] R. N. Braithwaite, "Pruning strategies for a volterra series model used in digital predistortion (dpd) of rf power amplifiers," in *2017 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, 2017, pp. 4–7.

[87] Y.-H. Kim, G. D. Jo, J.-H. Oh, J. H. Jung, J. H. Kim, C. Yu, and K. Lee, "An efficient simplified behavioral model for rf power amplifiers," in *2011 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications*, 2011, pp. 65–68.

[88] P. L. Gilabert, G. Montoro, T. Wang, M. N. Ruiz, and J. A. García, "Comparison of Model Order Reduction Techniques for Digital Predistortion of Power Amplifiers," in *2016 46th European Microwave Conference (EuMC)*, Oct 2016, pp. 182–185.

[89] Q. A. Pham, D. López-Bueno, G. Montoro, and P. L. Gilabert, "Dynamic selection and update of digital predistorter coefficients for power amplifier linearization," in *2019 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, 2019, pp. 1–4.

[90] Y. Li, X. Wang, and A. Zhu, "Sampling Rate Reduction for Digital Predistortion of Broadband RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 3, pp. 1054–1064, 2020.

[91] H. J. Newton, C. F. Baum, N. Beck, a. C. Cameron, D. Epstein, J. Hardin, B. Jann, S. Jenkins, and U. Kohler, "The Stata Journal," *Stata Journal*, vol. 10, pp. 288–308, 2010. [Online]. Available: http://ideas.repec.org/a/tsj/stataj/v7y2007i4p465-506.html

[92] P. L. Gilabert, G. Montoro, D. López, N. Bartzoudis, E. Bertran, M. Payaró, and A. Hourtane, "Order Reduction of Wideband Digital Predistorters Using Principal Component Analysis," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013, pp. 1–7.

[93] G. H. Golub and C. Reinsch, "Singular Value Decomposition and Least Squares Solutions," *Numer. Math.*, vol. 14, no. 5, pp. 403–420, Apr. 1970.

[94] A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[95] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[96] K. Finnerty, J. Dooley, and R. Farrell, "Utilizing Sparse-Aware Volterra for Power Amplifier Behavioral Modeling," *Proceedings from the 17th Research Colloquium on Communications and Radio Science into the 21st Century*, 2014.

[97] D. Wisell, J. Jalden, and P. Handel, "Behavioral Power Amplifier Modeling Using the LASSO," in *2008 IEEE Instrumentation and Measurement Technology Conference*, 2008, pp. 1864–1867.

[98] R. J. Tibshirani, "The Lasso Problem and Uniqueness," *Electronic Journal of Statistics*, vol. 7, no. none, pp. 1456 – 1490, 2013. [Online]. Available: https://doi.org/10.1214/13-EJS815

[99] H. Paaso and A. Mammela, "Comparison of Direct Learning and Indirect Learning Predistortion Architectures," in *2008 IEEE International Sympo-*

sium on Wireless Communication Systems, Oct 2008, pp. 309–313.

[100] K. Fyhn, T. L. Jensen, T. Larsen, and S. H. Jensen, "Compressive sensing for spread spectrum receivers," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2334–2343, 2013.

[101] Youngcheol Park, Wangmyong Woo, R. Raich, J. Stevenson Kenney, and G. T. Zhou, "Adaptive predistortion linearization of rf power amplifiers using lookup tables generated from subsampled data," in *Proceedings RAWCON 2002. 2002 IEEE Radio and Wireless Conference (Cat. No.02EX573)*, 2002, pp. 233–236.

[102] Z. Wang, J. Dooley, K. Finnerty, and R. Farrell, "Selection of compressed training data for rf power amplifier behavioral modeling," *Proceedings of the 10th European Microwave Integrated Circuits Conference*, 2015. [Online]. Available: http://mural.maynoothuniversity.ie/6447/

[103] Y. Li, X. Wang, and A. Zhu, "Sampling rate reduction for digital predistortion of broadband rf power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 3, pp. 1054–1064, 2020.

[104] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited volterra series-based digital predistortion for wideband rf power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 4198–4208, 2012.

[105] P. Desgreys, V. N. Manyam, K. Tchambake, D.-K. G. Pham, and C. Jabbour, "Wideband power amplifier predistortion: Trends, challenges and solutions," in *2017 IEEE 12th International Conference on ASIC (ASICON)*, 2017, pp. 100–103.

[106] Y. Beltagy, P. Mitran, and S. Boumaiza, "Direct learning algorithm for digital predistortion training using sub-nyquist intermediate frequency feedback signal," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 1, pp. 267–277, 2019.

[107] M. H. Hayes, *Statistical digital signal processing and modeling*. Wiley, 2014.

[108] E. Hsu, *An Essential Step to Generate a Modulation Signal – CCDF*, 2018. [Online]. Available: https://blogs.keysight.com/blogs/tech/rfmw.entry.html/2018/11/29/essential$_s$*teps*$_t$*og−ktGg.html*

[109] M. Garrido, P. Källström, M. Kumm, and O. Gustafsson, "Cordic ii: A new improved cordic algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 2, pp. 186–190, 2016.

[110] A. I. Dalbah, O. Hammi, and A. Zerguine, "Hybrid Look-Up-Tables Based Behavioral Model for Dynamic Nonlinear Power Amplifiers," *IEEE Access*, vol. 8, pp. 53 240–53 249, 2020.

[111] U. Cerasani, Y. Le Moullec, and T. Tong, "A practical FPGA-based LUT-predistortion technology for switch-mode power amplifier linearization," in *2009 NORCHIP*, 2009.

[112] L. Guan and A. Zhu, "Low-cost FPGA implementation of volterra series-based digital predistorter for RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 4, pp. 866–872, apr 2010.

[113] N. Mrabet, I. Mohammad, F. Mkadem, C. Rebai, and S. Boumaiza, "Optimized hardware for polynomial digital predistortion system implementation," in *RWW 2012 - Proceedings: 2012 IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications, PAWR 2012*, 2012, pp. 81–84.

[114] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 3, pp. 980–987, mar 2017.

[115] P. P. Campo, A. Brihuega, L. Anttila, M. Turunen, D. Korpi, M. Allen, and M. Valkama, "Gradient-Adaptive Spline-Interpolated LUT Methods for Low-

Complexity Digital Predistortion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 1, pp. 336–349, jan 2021.

[116] S. Liang, Z. Jiang, L. Qiao, X. Lu, and N. Chi, "Faster-Than-Nyquist Precoded CAP Modulation Visible Light Communication System Based on Nonlinear Weighted Look-Up Table Predistortion," *IEEE Photonics Journal*, vol. 10, no. 1, feb 2018.

[117] S. Zhalehpour, J. Lin, H. Sepehrian, W. Shi, and L. Rusch, "Experimental Demonstration of Reduced-Size LUT Predistortion for 256QAM SiP Transmitter," in *2019 Optical Fiber Communications Conference and Exhibition (OFC)*. San Diego, CA, USA: IEEE, 2019.

[118] N. Naraharisetti, P. Roblin, C. Quindroit, and S. Gheitanchi, "Efficient least-squares 2-d-cubic spline for concurrent dual-band systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 7, pp. 2199–2210, 2015.

[119] R. Zitouni and L. George, "Output power analysis of a software defined radio device," in *2016 IEEE Radio and Antenna Days of the Indian Ocean (RADIO)*, 2016, pp. 1–2.