2022-08

Digital Pre-Distortion for Radio Transmitters with Multiple Sources of Impairment

Motaqi, Ahmadreza

UNIVERSITY OF CALGARY

Digital Pre-Distortion for Radio Transmitters with Multiple Sources of Impairment

by

Ahmadreza Motaqi

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ELECTRICAL AND COMPUTER ENGINEERING

CALGARY, ALBERTA

AUGUST, 2022

Abstract

Among the most important factors in wireless base-stations design are link speed and power efficiency. For a faster wireless connection, the MIMO beamforming technology, higher-order QAM signals, and wider bandwidth signals are used. To increase the power efficiency in transmitters, the PAs are pushed to operate in nonlinear regions where they present their highest power efficiency. These techniques increase in-band and out-of-band distortions of the MIMO wireless transmitters, resulting in signal quality degradation. To mitigate the introduced distortions to the transmitted signal, Digital Pre-Distortion (DPD) is used. However, the conventional DPD techniques cannot meet the strict timing requirements of 5G and 6G wireless connections. The transmitter nonlinearities are a function of various parameters such as input signal average power and ambient temperature and steering angle in the case of beamforming.

This thesis studies the effect of ambient temperature and signal's average power on high-power PAs and proposes a novel DPD technique to mitigate the effect of those parameters on the PA behaviour. In another research, the effect of beamforming on the transmitter performance in terms of signal quality and out-of-band distortions are studied. A novel angle inclusive DPD for beamforming application is introduced to remove beamforming-related distortions and enhance the signal quality. This technique provides an uninterrupted linearization at any beam direction. The developed algorithms are verified using a realistic MIMO beamforming setup designed in the laboratory. The measurement results have shown that by using the proposed technique, the signal quality is substantially improved when compared to state-of-the-art techniques.

Acknowledgments

Dedication

I would like to dedicate this thesis to myself for all the work I have done, to myself, a person

who has never given up on his **dreams** despite all the challenges and hardships…

To my dad Dr. Ali Motaqi and to the best mom, Zahra who taught me discipline

Table of Contents

List of Figures

List of Tables

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
ACPR	Adjacent Channel Power Ratio
ADC	Analog-to-Digital Converter
AI	Artificial Intelligence
AI	Angle Inclusive
P	Number of Volterra series kernels
DAC	Digital to Analog Converter
DPD	Digital Pre-Distortion
5G	Fifth generation of internet connection
GHz	Giga Hertz
$a_{k,m}$	Memory polynomial coefficient for the kth nonlinearity and m^{th}
PLL	Phased Locked Loop
LO	Local Oscillator
SDR	Software Defined Radio
NMSE	Normalized Mean Square Error
PAE	Power Added Efficiency
PA_{RFout}	PA RF output
PA_{RFin}	PA RF input
PA_{DC}	The input power consumption of the paper
£	Amount of IQ imbalance for the complex envelope of the input
¥	Amount of IQ imbalance for conjugate of the complex envelope of the input
$x_{out-mixer}$	Out put of mizer
x	Complex Input signal

x^*	Conjugate of the complex Input signal
I	Real part of the complex input
Q	Imaginary part of the complex input
M	Memory taps of the memory polynomial
K	Nonlinearity order of the memory polynomial
x_i	Inphase part of the input signal
x_q	quadrature part of the input signal
$\Delta\widetilde{\emptyset}$	Phase difference between two complex signals
$\text{\euro}_1$	Phase offset of signal 1
λ	the signal wavelength,
ϕ	the elevation angle
θ	the azimuth angle
d_x	distances between the adjacent elements of the phased array antenna in x-axis
d_y	distances between the adjacent elements of the phased array antenna in y-axis
AGC	Adaptive Gain
LNA	Low Noise Amplifier
MIMO	Multi input multi output
SISO	Single input single output
SDR	Software defined radio
FPGA	Field Programable Gate Arrays
Γ	coefficient reflection
$a_{m_1,k_1}(\Gamma)$	reflection dependant model coefficients with k_1 nonlinearity other and m_1 memory depth
CNN	Convolutional neural network
FCNN	Fully connected neural network

Symbol	Description
w_s	the beamforming weights for linear phased array antenna
P	the number of radiating elements in linear phased array antenna
$J_k(z)$	kth Bessel function
d^{th}	Filter number in CNN model
CK_out_d	output of the dth filter
$W_{in} \times L_{in}$	Size of the input image to the CNN model
f_d	Activation function in the dth filter
$Bias\_filter_d$	Bias in the dth activation function in convolutional layer
$Image_i$	Input image to the CNN model
$Filter_d$	The dth Filter coefficients
$W_{i,j}^{d,g}$	Connecting weight between the (i,j) element of the dth filter volume and g^{th} neuron of the fully connected neural network layer
$W_{1,g}$	connecting weights between the gth neuron in the first layer and output neuron number one.
$W_{2,g}$	connecting weights between the gth neuron in the first layer and output neuron number two.
$Bias\_FC_g^1$	bias corresponding to the gth neuron in the first NN layer
$Bias\_FC_1^2$	biases for the first neuron in the output layer of the NN
$Bias\_FC_2^2$	biases for the second neuron in the output layer of the NN
$FC\_out_g^1$	output of the gth neuron in the first layer of the fully connected layer
G	Number of the neurons in the first layer of the fully connected neural network

MSE(r)	Mean Square Error at iteration r
F	the number of filters in a single convolutional layer
L_F	length of the convolution filters
W_F	width of the convolution filters
$padding_{rows}$	
$padding_{column}$	
FC	Fully connected layerL_F
$multiplications_{CK}$	The number of multiplications in convolutional layer
$addition_{CK}$	the number of addition operation
AF_{ck}	the number of FLOP required for activation functions
$multiplications_{FC}$	number of multiplications in the fully connected layer
$addition_{FC}$	number of additions in the fully connected layer
N_1	Number of neurons in the first layer of fully connected neural network
N_2	Number of neurons in the second layer of fully connected neural network
Q	The total number of filter coefficients
Z_{FC}	The total number of learnable parameters in the FC layer
H	the number of required FLOPs per activation function.
PSD	Power Spectral Density

1 Chapter 1: Motivations and Introduction

1.1 Motivations and Introduction

The need for higher-speed wireless communication systems has led researchers and scientists to develop new techniques to increase link speed and reliability. One way to increase the connection speed is to increase the spectral efficiency, i.e., sending more bits per hertz. To increase spectral efficiency, different modulation schemes like Orthogonal Frequency Division Multiplexing (OFDM) are used in WiMAX and Long-Term Evolution (LTE) applications, and Filtered Orthogonal Frequency Division Multiplexing (FOFDM) for in 5th Generation (5G) applications are developed [1]. Using higher-order Quadrature Amplitude Modulation (QAM) in 5G applications also improves the connection speed; however, it increases the Peak-to-Average Power Ratio (PAPR). High PAPR reduces the Power Amplifier's (PA) power consumption efficiency in wireless transmitters. To mitigate the PA nonlinearities, Digital Pre-distortion (DPD) is used. DPD is the inverse model of the nonlinearities in such a way that when it is cascaded with the PA, the output of the PA has a linear relation with the input signal. This improves the signal quality and reduces out-of-band distortions.

To further increase the spectral efficiency, beamforming transmitters are developed. Beamforming transmitters steer the beam toward the users located at different special locations. Beamforming and Multi Input Multi Output (MIMO) transmitters are prone to be noisier and more distorted since crosstalk and coupling effects exist between the RF channels which degrade the signal quality. The DPD design for MIMO beamforming transmitter is more challenging compared to Single Input Single Output (SISO) DPDs as the nonlinearities and distortion are direction dependant i.e. varying versus steering angle. The DPD convergence time is another bottleneck in MIMO beamforming DPD design. The linearizer should be able to achieve the required

linearization performance within a limited time span (~ 1ms). Moreover, the base stations are supposed to operate at different environmental locations at different transmitting power. The environmental and transmitter setting parameters also influence the transmitter behavior, making them vital to be considered in DPD design. These challenges make the beamforming DPD design more difficult compared to the SISO DPDs.

1.2   Research Objectives and Contributions

Considering the challenges ahead of beamforming DPD design, this thesis focuses on introducing a smart DPD technique for both SISO and MIMO beamforming transmitters for 5G applications. More specifically, this work aims for three main objectives listed as follows:

1. **MIMO beamforming wireless transmitter platform design procedure**

    To validate a beamforming DPD algorithm for practical applications, a reliable beamforming setup is essential. This thesis provides a detailed procedure to design and implement a MIMO beamforming platform using off-the-shelf products. Moreover, a systematic procedure to calibrate and test the MIMO platform is proposed.

2. **Angle Inclusive Beamforming DPD for phased array transmitters**

    Direction dependency of the distortions in beamforming transmitters requires the DPD coefficients to be updated at each steering direction. DPD identification is time-consuming, and the transmitter might not be able to meet the timing requirements. 5G wireless transmitters are supposed to have lower than 1ms latency. This latency is the time between transmitting and receiving information. Therefore, the DPD convergence time need to fast

so that it transmits high quality signal while meeting the ACPR mask for the operating frequency [51]. This thesis proposes an Angle Inclusive DPD (AI-DPD) architecture that includes the beam steering direction in the DPD model. The AI-DPD provides uninterrupted linearization across the steering range and reduces the hardware complexity by eliminating the need for a dedicated linearization performance for each subarray.

3.   **Effect of ambient temperature and signal's average on the PA's performance**

Wireless base stations are supposed to operate in different environments with a large span of temperature variation. The Signal's average power also affects the PA's junction temperature and alters the behavior of the PA. This thesis studies the effect of both ambient temperature and signal's average power on the PA's behavior jointly and proposes a DPD technique that includes both factors in the DPD model. The proposed Power-Temperature Inclusive DPD (PTI-DPD) provides uninterrupted linearization performance at any combination of DPD parameters and reduces the training complexity as it requires few training conditions.

1.3   Thesis outline

Chapter II studies the state-of-the-arts DPD techniques for both SISO and MIMO beamforming applications. This chapter provides a deep insight into work that has been done by other researchers and studies the advantages and disadvantages of these techniques.

Chapter III provides the design, implementation, and calibration details of MIMO phased array antenna transmitters. The platform design is described in three sections. The first section discusses the digital tools and the FPGA design for MIMO beamforming platforms. Then, system interconnections are discussed. Section 2 discusses the frequency coherence and phased array antenna and provides a calibration procedure for MIMO wireless systems. This section discusses the calibration to remove any DC offset, IQ imbalances, and any phase variation in RF channels. Moreover, a validation test is proposed to observe the quality of the calibration.

Chapter IV proposes an angle-inclusive DPD for beamforming applications using the Convolutional Neural Network (CNN). For the NN or CNN model to be able to provide the best modeling performance, as well as lower complexity, it should be fed with the right information. This chapter studies the important parameters affecting the beamforming transmitter and proposes an input image with pre-processed data to increase the modeling performance.

Chapter V studies the effect of the ambient temperature and the signal's average power on the PA performance. The performance of a high-power Doherty PA under different ambient temperatures and signal's average power has been studied. It has been observed that the ambient temperature and the signal's average power variation have a destructive effect on the linearization performance. To address this issue, a novel PTI-DPD is developed to compensate for the distortions under varying ambient temperature and the signal's average power conditions.

The final chapter summarizes this thesis and discusses future works and recommendations.

2 Chapter 2: Literature Review

2.1   Background and SISO DPDs

Power amplifiers are the most power-hungry parts of wireless systems and it is important to operate them in high-energy efficiency regions. Figure 2.1 shows the Power Amplifier Added Efficiency (PAE) of a custom PA as a function of input power. The PAE is calculated as follows:

$$PAE = \frac{PA_{RFout} - PA_{RFin}}{P_{DC}} \tag{2.1}$$

Here the PA_{RFout} and PA_{RFin} are the RF output power and RF input power, respectively. The P_{DC} is the drawn DC power from the power source. The closer to the saturation region (higher output power), the higher the efficiency of a PA. Higher PAPR signals force average power to operate at power back-off regions with low power efficiency. To increase the power efficiency, the PA is pushed to operate closer to saturation regions. However, the PA presents a highly nonlinear behaviour in these regions.
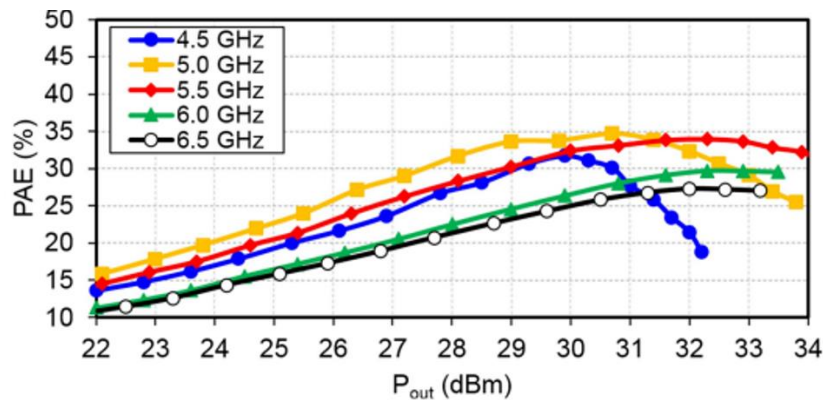
©2021 IEEE

Figure 2.1 Power Added Efficiency of a custom designed PA versus output power at different frequencies [2].

The nonlinear behaviour of the PA reduces the signal quality presented in the term of Normalized Mean Square Error (NMSE) and increases the Adjacent Channel Power Ratio (ACPR). The NMSE and ACPR are calculated as follows:

$$NMSE = 10.log\left(\frac{\frac{1}{N}\sum_{n=1}^{n=N}\left[\left(I(n)-I'(n)\right)^2+\left(Q(n)-Q'(n)\right)^2\right]}{\frac{1}{N}\sum_{n=1}^{n=N}\left[\left(I(n)\right)^2+\left(Q(n)\right)^2\right]}\right) \qquad (2.2)$$

$$ACPR = \frac{Power\ in\ the\ adjacent\ channel}{Power\ in\ the\ main\ channel} \qquad (2.3)$$

Where N is the number of samples. I', Q', I and Q are the modeled and measured I and Q data, respectively.

The ACPR and signal quality are very important, and they are determined by the regulatory organizations. Higher ACPR causes interferences with adjacent channels and reduces the sensitivity or may even block other receivers. Therefore, it is crucial for a transmitter to meet the ACPR standards.

In order to understand the effect of nonlinear behaviour of the PA on the ACPR of the transmitter, let us assume the PA output can be described using 5 order polynomials as follows:

$$P_{out} = a_0 + a_1.x + a_2.x^2 + a_3.x^3 + a_4.x^4 + a_5.x^5 \qquad (2.4)$$

Where the a_n is the polynomial coefficient for n^{th} order term and x is the input signal to the PA.

And input x is:

$$x = Ae^{(w_1.t+\rho_1)j} + Be^{(-w_2.t+\rho_2)} \qquad (2.5)$$

Using two tone excitation technique, the PA output inter-modulation products can be predicted. The P_{out} contains the frequencies presented in TABLE 2.1. The main culprits for increased ACPR in the adjacent channels are the $2w_1 \mp w_2$, $w_1 \mp 2w_2$ called third order inter-modulations and for the raised ACPR in alternate channel are $3w_1 \mp 2w_2$, $2w_1 \mp 3w_2$ called fifth order inter-modulations. Generally speaking, odd order inter-modulations are closer to the main frequency band, more difficult to filter out and they increase the ACPR.

Table 2-1 Inter-modulation products using two tone excitation signals

Nonlinearity order	Inter-modulations frequency		
First order	w_1, w_2	-	-
Second order	$2w_1, 2w_2$	$w_1 \mp w_2$	-
Third order	$3w_1, 3w_2$	$2w_1 \mp w_2$ $w_1 \mp 2w_2$	-
Forth order	$4w_1, 4w_2$	$2w_1 \mp 2 w_2$	$w_1 \mp 3w_2$ $3w_1 \mp w_2$
Fifth order	$5w_1, 5w_2$	$4w_1 \mp w_2$ $w_1 \mp 4w_2$	$3w_1 \mp 2w_2$ $2w_1 \mp 3w_2$

To compensate for the introduced nonlinearity to the transmitted signal, Digital Pre-Distortion (DPD) technique is used [2]. Figure 2.2 illustrates the basic operation of the DPD. DPD is the inverse model of the transmitter nonlinearities in way that when it is cascaded with the transmitter, the final output has a linear relation compared to the input signal. Using DPD technique, the PA can operate in higher efficiency regions with a good linearization performance across the steering range.
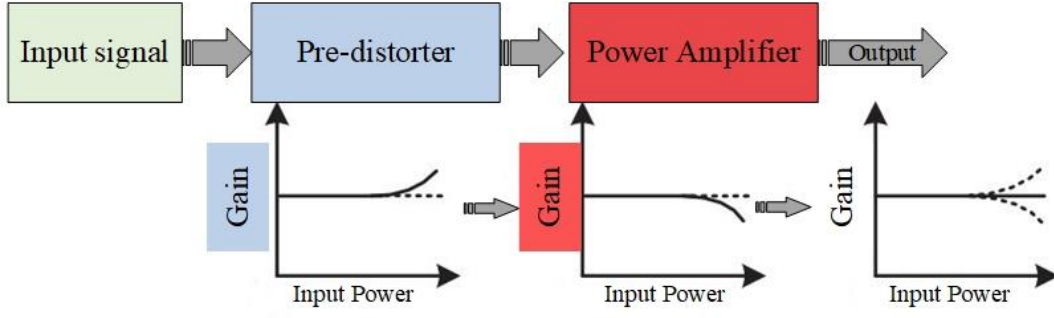
Figure 2.2 Basic DPD concept for a SISO system

2.1.1 State-of-the-art SISO DPD

So far, different DPD techniques are proposed for Single Input Single Output (SISO) Radio Frequency (RF) transmitters. Look-Up Table (LUT) DPD is the simplest architecture to linearize a wireless transmitter. In this technique, the signal power is quantized into N points and a complex number is multiplied to signal in each level [4]. Volterra series [5] is the most accurate and complex way of modeling the PA behaviour. Equation (2.6) presents the estimated PA output y using p^{th} Volterra Kernels order h_p, applied to input signal $x(n)$ with m_p memory depth.

$$y(n) = h_0 + \sum_{p=1}^{p=P} \sum_{m_1=a}^{m_1=b} \dots \sum_{m_p=a}^{m_p=b} h_p(m_1, m_2, \dots m_p) . \prod_{k=1}^{k=p} x(n - m_k) \qquad (2.6)$$

Due to the excessive complexity of Volterra series, it is not a suitable model for practical application. Memory polynomial is a simplified version of Volterra series which is widely used for PA modeling and DPD design [6]. The output a PA modeled using memory polynomial is calculated as follow:

$$y(n) = \sum_{m=0}^{M-1} \sum_{k=1}^{K-1} a_{k,m} \cdot x(n - m) \cdot |x(n - m)|^{k-1} \qquad (2.7)$$
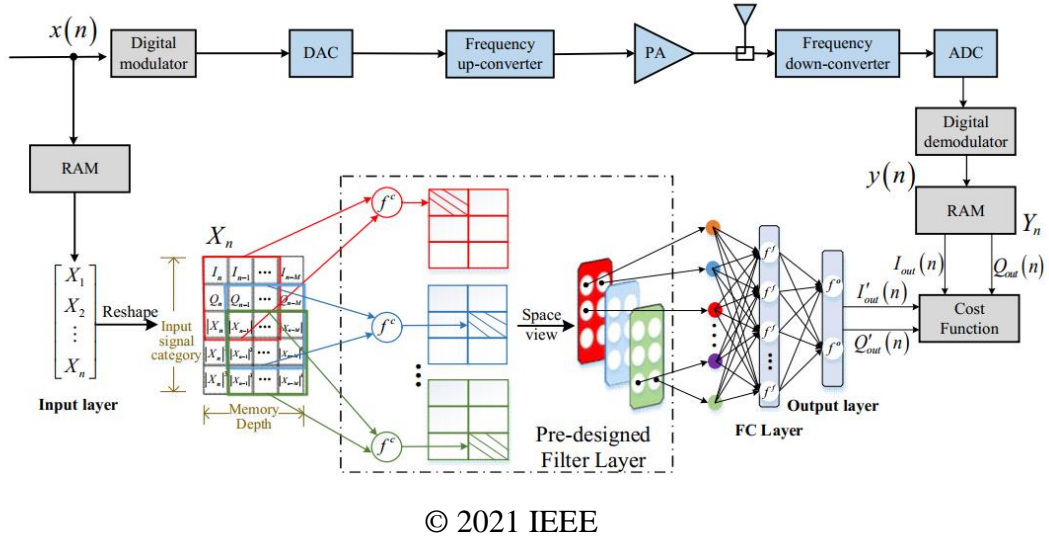
Figure 2.3 The architecture of the developed DPD for wideband signal linearization [11]

Here the $y(n)$ and $x(n)$ are the PA output and input samples respectively. The $a_{k,m}$ is the coefficient corresponding to the mth memory depth and k^{th} nonlinearity order. A more accurate way to describe the PA behaviour is using Generalized memory polynomial. The Generalized memory polynomial is an augmented version of memory polynomial with cross-terms that increases the modeling accuracy. Generalized memory polynomial has been widely used for DPD architecture in both academy [7] and industry. Some other approaches adopt more advanced techniques like Neural Network (NN) [8][9][10] or Convolutional Neural Network (CNN) [11]. As illustrated in Figure 2.3, the authors in [11] have used the CNN to linearize wideband signals. The input IQ signal with M delays and K nonlinearity order are organized into a 2-D array, and used to train the DPD model. The CNN is composed of layer of convolutional layer cascaded with a shallow neural network to generate the DPD signal. They have shown the proposed DPD in [11] is able to show a better performance in comparison to the other techniques. However, their technique is a very primitive application of the CNN and they were not able to benefit from all

aspects of the CNN. The authors have not provided the DPD model with information about the signal or PA conditions; therefore, all the feature extraction burdens are on the CNN model which substantially increases the model complexity. In similar research in [12], the CNN is used to train a direct and indirect learning DPD to linearize high baud-rate signals transmitted over fiber. They have used three layers of CNN to pre-distort the signal as it is shown in Figure 2.4. The section A and C shown in Figure 2.4 are the convolving layers which are linear operations, and the section B is a fully connected NN which models the nonlinear behaviour of the system. They have shown how using the architecture shown in Figure 2.4 substantially improves the performance in
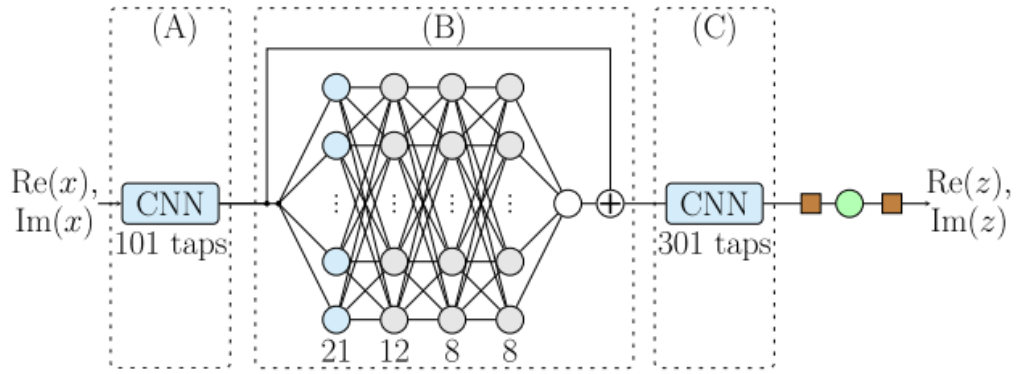
Figure 2.4 The architecture of the proposed CNN-DPD in [12].

comparison to other techniques.

Figure 2.5 shows the architecture of the DNN SISO DPD developed in [10]. The inputs to the system are the delayed version of the modulated signal. The main disadvantage of these techniques is that their models do not consider other important factors like signal's average power, reflection coefficients, and ambient temperature that manipulate the PA behavior. To address this issue, the authors in [13] have proposed a reflection coefficient-aware DPD model using a cross-over [14]
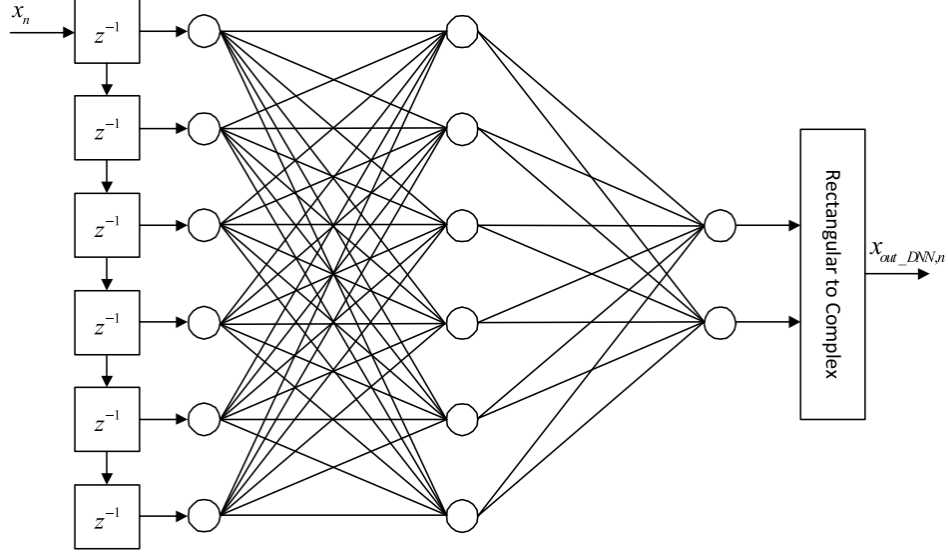
© 2020 IEEE

Figure 2.5 Deep Neural Network DPD for SISO applications [10].

memory polynomial. In this technique, the DPD coefficients are updated based on the reflection coefficients under mutual coupling. The output $y(n)$ of the first PA in the presence of cross-talk due to input signal x_1 and x_2 of is calculated as follows [13]:

$$y(n) = \sum_{m_1=0}^{M_1} \sum_{k_1=0}^{K_1} a_{m_1,p_1}(\Gamma).x_1(n-m_1).|x_1(n-m_1)|^{k_1} \times |x_2(n-m_1)|^{k_1}$$

$$+ \sum_{m_2=0}^{M_2} \sum_{k=0}^{K_2} a_{m_2,p_2}(\Gamma).x_2(n-m_2).|x_2(n-m_2)|^{k_2} \times |x_1(n-m_2)|^{k_2} \ (2.7)$$

Where Γ is the coefficient reflection, $a_{m_1,k_1}(\Gamma)$ denotes the reflection dependant model coefficients with k_1 nonlinearity other and m_1 memory depth. They have observed that the reflection-aware DPD is able to significantly improve the ACPR and NMSE of the error between the original and received signal.

24

The PAs behaviour is also a function of signal's average power. A power adaptive DPD model based on memory polynomial is proposed in [15] where the model coefficients are selected based on the signal's average power. Figure 2.6 illustrates the principle of the [15] model operation. In this technique, a reference DPD is extracted at a certain power in parallel to linear and nonlinear corrections to model the PA at different power levels.
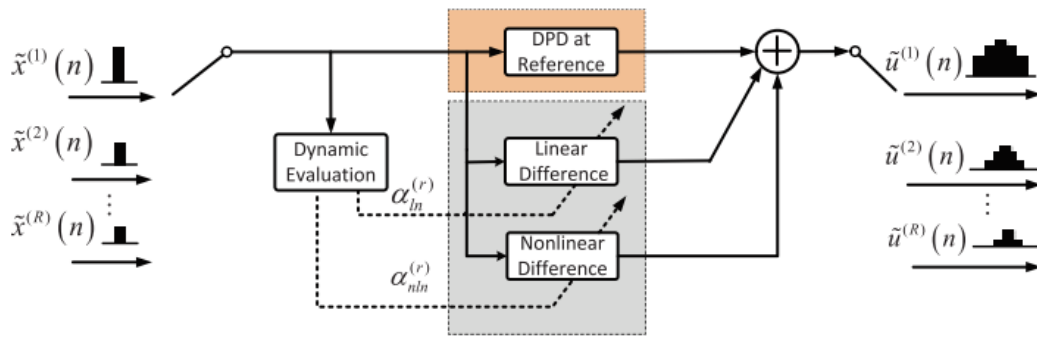
Figure 2.6 The principle of the power adaptive model operation [15].

In another work [16], a two-box parallel DPD architecture is proposed to mitigate the effect of signal's average power. Figure 2.7 shows the architecture of the proposed technique in [16]. They have shown that PA behavior can be classified into static and dynamic behavior. The dynamic behaviour is a function of signal's bandwidth, and it is modeled using a memory polynomial. The static behaviour is a function signal's average power, and a LUT is provided to compensate for any variation in PA behaviour due to signal's average power.
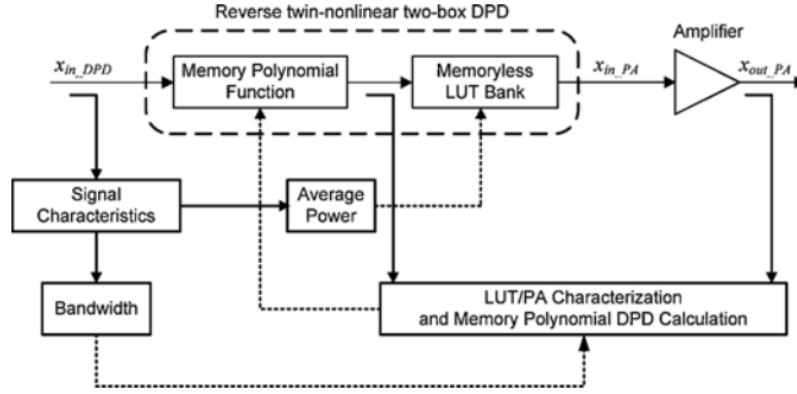
© 2013 IEEE

Figure 2.7 The architecture of the proposed power scalable DPD in [16].

In very recent research in [17] (this paper has been published contemporary to the work done in this thesis), the authors have proposed an intelligent sensing technique that considers the signal's average power, bandwidth and modulation scheme to pre-distort the signal accordingly. To reduce the complexity of the neural network, a self-sensing module is designed. The self-sensing module is composed of different functions extracting bandwidth, signal power and PAPR of the signal. The outputs of the self-sensing module are used to model the proposed DPD architecture [17]. Moreover, the model in [17] contains static polynomial block to capture the static features. The architecture of the proposed DPD is shown in Figure 2.8. The final output of the model is a NN fed with inputs from self-sensing module, static polynomial kernels and the input modulated signal.
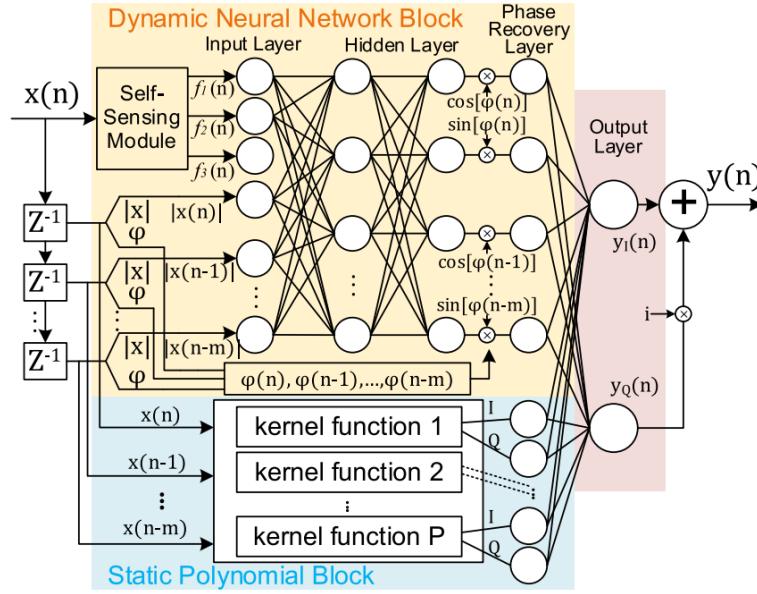
Figure 2.8 The architecture of self-sensing DPD [17].

In addition to the aforementioned factors affecting the PA behaviour, there are other environmental parameters such as ambient temperature. The cellular base stations are supposed to operate in environments where the ambient temperature can range from -40° Celsius to +50°. To maintain the signal quality, the effect of ambient temperature on the PA behaviour should be compensated. So far, no existing DPD architecture has considered the ambient temperature in their DPD architecture.

The previously established DPD architectures consider up to one important factor affecting the PA's behavior (i.e. input signal and one factor) and they ignore the rest of the parameters. This thesis bridges this gap and introduces a novel DPD model that includes the modulated signal, ambient temperature, and signal's average power to pre-distort the signal using Artificial Intelligence (AI).

27

2.2   MIMO beamforming DPD

The upcoming generations of internet connections are expected to support an enormous number of users with increasing demand for faster connection speed. Considering the limitations in the available frequency spectrum, Multi Input Multi Output (MIMO) beamforming transmitters are developed for further improvement in spectral efficiency. As shown in Figure 2.9, in MIMO beamforming the same frequency band can be used, but in different directions. The narrower the steered beam, the more the number of supported users in the same frequency band and the greater the spectral efficiency. Beamforming transmitters benefit from phased array antennas to steer the beam toward the users. Generally, there are three types of beamforming: digital beamforming, analog beamforming, and hybrid beamforming [18]. In digital beamforming, the weights of antenna elements are applied in the digital domain to the base-band signal. Digital beamforming is more accurate and has been deployed in many satellite applications [19], but it is more expensive and bulkier compared to the other techniques. However, it is suitable for satellite and military applications. Analog beamforming uses lens or analog phase shifters to concentrate the beam toward the users [20]. The hybrid beamforming uses digitally controlled analog phase shifters [21]. Due to quantization error in analog phase shifters, there is always an error in hybrid beamformers. Nevertheless, hybrid beamformers are suitable for commercial massive MIMO beamforming transmitters as they reduce the hardware complexity.
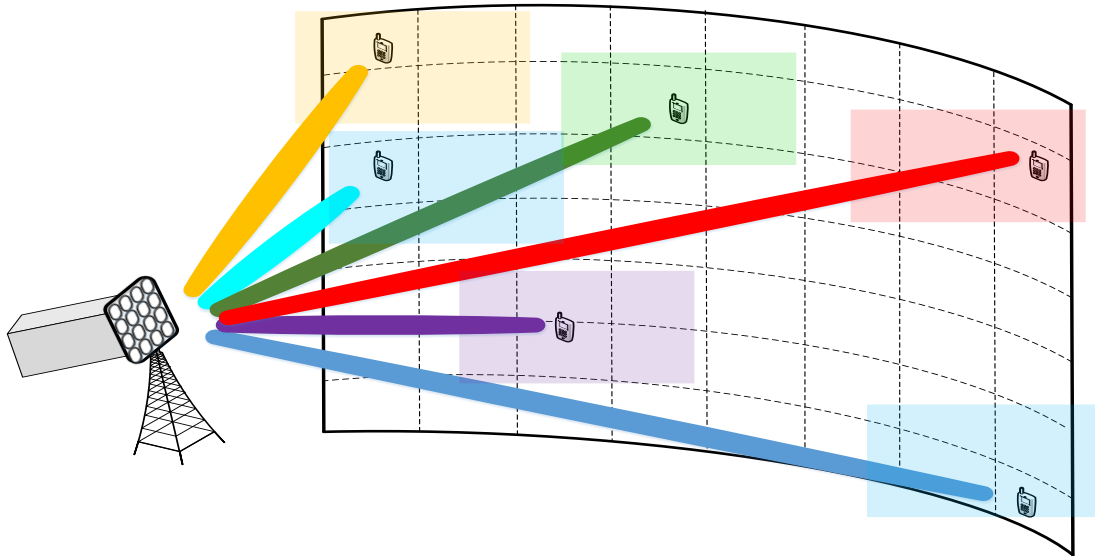
Figure 2.9 Operation of MIMO beamforming wireless transmitter.

In addition to the aforementioned challenges in SISO wireless transmitters, by increasing the number of RF channels in the MIMO beamforming, the crosstalk and coupling effect between the antenna elements increases. Coupling effect is the electromagnetic leakage of the adjacent RF channels to other channels. The coupling effect changes impedance matching of the antenna and causes a reflection in the RF transmission path, thereby degrading the signal quality. More importantly, the coupling effect is a function of steering angle, and it varies as the beam is steered toward the different directions [22]. As depicted in Figure 2.10, if the crosstalk happens before the nonlinear component like PA, it is nonlinear. If crosstalk happens after PA, between antennae, it is linear crosstalk. The more compact the antenna elements, the greater the coupling effect. In
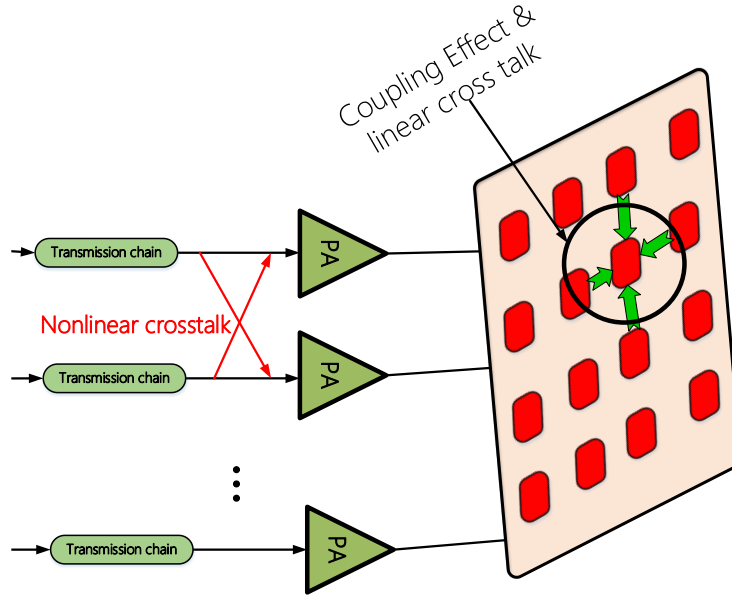
Figure 2.10 coupling effect between the antenna elements, nonlinear and linear crosstalk.

industrial applications, the RF front ends target a trade-off between the antenna size and the coupling effect.

There are many challenges ahead of deploying massive MIMO beamforming DPDs for 5G applications. The first challenge is the strict timing requirement for the DPD to converge. Radio base stations are supposed to operate at variable average power, ambient temperature, different steering direction and other varying conditions. At different conditions, the transmitter's behavior changes and the DPD coefficient should be updated. For 5G applications, the latency between the transmitted signal and received one is expected to be 1 milliseconds [23]. This latency is the time between transmitting and receiving information. Therefore, the DPD convergence time-DPD reidentification time- need to fast so that it transmits high quality signal while meeting the ACPR mask for the operating frequency [23]. Such a strict timing requirement is difficult to satisfy. The other challenge is hardware complexity; having one dedicated observation path per subarray increases the hardware complexity, power consumption, and price of 5G cellular base stations.

Many research groups have focused on introducing new techniques to address the linearization challenges for wireless transmitters. The following section studies the relevant state-of-the-art beamforming DPD techniques.

2.2.1 State-of-the-art MIMO beamforming DPDs

The DPD design for beamforming and MIMO transmitter is even more challenging than SISO DPD. As mentioned earlier, the direction dependency of the distortion, strict timing requirements for 5G, and the hardware complexity are the main challenges ahead of MIMO beamforming DPD design. The state-of-the-arts can be studied from two perspectives: feedback signal acquisition and beamforming DPD architecture.

As depicted in Figure 2.11, three techniques have been proposed to acquire the feedback signal. The first one shown in Figure 2.11 (a), is the far-field signal which is commonly used among researchers [24]. The second approach is the embedded receiver feedback path proposed in [25]. In the architecture shown in Figure 2.11 (b), the feedback paths are four antenna elements embedded in the antenna array operating a in time multiplexing method to collect data. They have shown that a symmetric receiver antenna around the Tx array can provide enough data to reconstruct the main beam. The authors have used the fact that coupling effect between antennas can be expressed as amplitude and phase. In the first step, the coupling model is extracted and used to estimate the far-field signal using the feedback antenna from the radiated signal of each TX path. Using this technique and repeating the processes for all antenna elements, while eliminating the uncorrelated data (coming from other subarrays), the main beam is estimated. The authors assume the identical behaviour of PAs in each channel; this assumption is inaccurate as the RF chains do not behave similarly. Moreover, they have assumed that the signal is repeated to reconstruct the beam throughout the observed channels, while in base stations, the signal is not

31

repeated. Such objections question the credibility of the proposed method in [25] for commercial

applications. The third feedback signal acquisition is using near-field fixed antenna [26] shown in

Figure 2.11 (c). The authors in [26] have used a fixed antenna element to capture the main beam
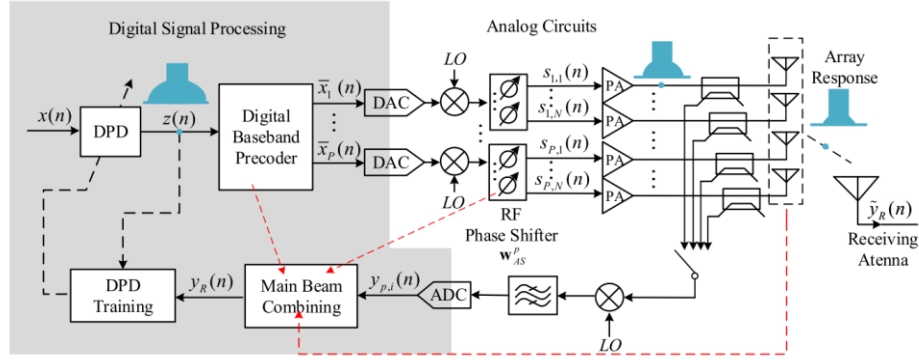
© 2018 IEEE

(a)
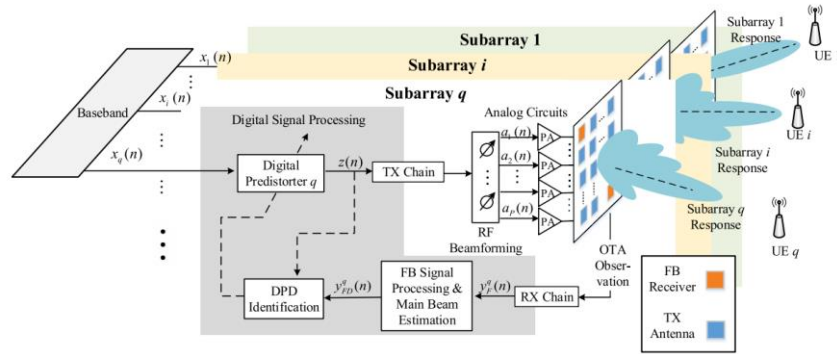
© 2019 IEEE

(b)
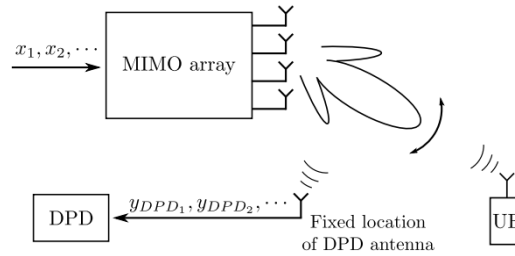
© 2019 IEEE

(c)

Figure 2.11.feedback signal acquisition using (a) far-field [24]; (b) near field embedded antenna
elements [25]; (c) near field fixed antenna [26].

and characterize the PAs. They have tried to simplify their technique to make it more practical by eliminating the need for couplers or switches while providing continuous transmission of data. It has been observed that in practical scenarios, for specific beam direction, the feedback signal can have a very low dynamic range and consequently, most of the nonlinear information falls below the noise floor. As a result, the DPD can not compensate for them. Therefore, using a fixed antenna for DPD identification is not able to linearize the beam across the whole steering range.

The state-of-the-art beamforming DPDs can be divided into three classes of architecture. The following paragraphs study each DPD architecture and discuss the advantages and disadvantages of each technique.

A.   Beam oriented Bo-DPD [24]: the far-field received signal is used to characterize the beamforming transmitter in the Bo-DPD. Figure 2.11 (a) and (b) depict the architecture of the Bo-DPD. Here, instead of linearizing each PA, the virtual far-field signal toward the desired direction is constructed and used to model the transmitter. The virtual beam is the sum of the PAs' signals in each sub-array. In Figure 2.11 (a) each RF channel is connected to couplers to monitor the signal. An RF observation path transfers the output of couplers to the PC in a time-multiplexing manner. This approach is not practical for applications where the number of RF channels is large. The second approach uses the embedded feedback receivers as shown in Figure 2.11 (b). Here, the couplers are removed but the observation path continues to work in a time-multiplexing manner to capture the output of RF channels. Equation (8) estimates the far-field virtual beam using the feedback signal through coupler [24] or embedded receivers [25].

$$y_\theta\left(n\right) = \sum_{i=1}^{i=I} e^{j(i-1).\psi} \times \sum_{m=0}^{M-1} \sum_{k=1}^{K-1} a_{k,m} \cdot x(n-m) \cdot |x(n-m)|^{k-1} \quad (2.8)$$

33

Here ψ is a constant phase offset added to each RF channel for beamforming purposes, and the rest of the terms present a memory polynomial describing the PA behaviour in each RF channel. For the sake of simplicity, the PAs in RF channels are considered identical. In the case of MIMO beamforming transmitters, they have suggested an iterative approach to remove the uncorrelated signal from the observation path. Then, a memory polynomial is extracted to linearize the virtual beam at each direction. It has been observed that the beam behavior varies versus steering angle. This is mainly due to coupling effect variation related to beamforming weights. Therefore, the DPD coefficients need to be updated as the beam is steered toward the different directions. This requires the DPD identification to be fast enough to meet the 5G timing requirements. Moreover, Bo-DPD required an independent observation path for each subarray to monitor the transmitted signal for DPD coefficient re-identification as the beam is steered toward different directions. This makes the base station system bulkier and more expensive, and it is not practical for massive MIMO applications where a limited number of observation paths are available.

B. Full angle DPD [27]: this technique is only applicable to a fully digital beamforming transmitter where for each RF channel, there is a digital path. Figure 2.12 shows the architecture of the proposed full angle DPD in [27]. In this technique, each PA is linearized individually by capturing the output signal from each channel. The DPD model is a two-box cascade DPD where a common DPD is extracted to model the common features and one fine-tuning box is considered per digital/RF channel to compensate for the difference in PA behaviour in each RF channel. The proposed
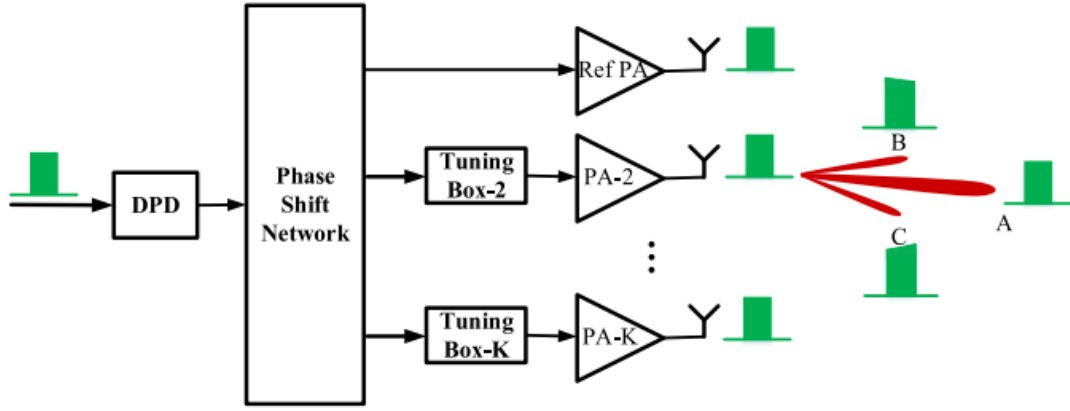
Figure 2.12 The architecture of the full angle DPD proposed in [27].

technique presents an acceptable linearization (≈20 dBc improvement) across the steering range without the need to update the DPD coefficients. However, this architecture requires a fully digital beamformer and a dedicated observation path for each element which makes this approach impractical for industrial applications. Moreover, the reason that the proposed technique in [27] is able to provide full range linearization is that the coupling between antenna elements are reported to be -19 dB which is too low to alter the behavior of the PA; while in practical scenarios, the coupling between antennae is stronger.

C. Neural Network-based beamforming DPD [28]: Figure 2.13 shows the architecture of the DNN-DPD proposed in [28]. Here, the far-field sum of the PAs in each subarray is acquired to estimate the weight of a Deep Neural Network (DNN) that models the inverse behaviour of the transmitter. The DNN is trained using 17 angles across the steering range and generates a single pre-distorted signal that linearizes the signal across the steering range. The DNN-DPD is trained to minimize the Mean Square Error (MSE) over the steering range, and as a result, it can not provide the
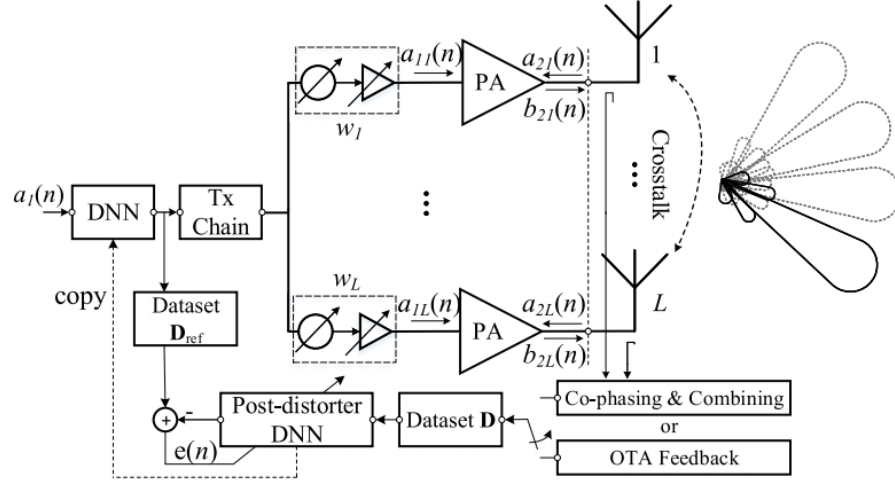
Figure 2.13 DNN-DPD for active array antenna linearization [28].

maximum achievable performance at each steering range. As the PA's behaviour varies during the beamforming, the pre-distorted signal at each direction must be different from the other directions, otherwise, the maximum linearization is not achieved. This approach is very similar to SISO DNN [9][10] or CNN DPD [11] and the only difference is that the [28] is trained over 17 sets of data.

The previously established beamforming techniques do not provide a practical DPD solution for 5G and other upcoming generations of internet connections as they either violate the timing requirements or increase the transmitter hardware complexity. This thesis addresses these issues by introducing a novel AI-based MIMO beamforming DPD that provides a low latency DPD and reduces the transmitter's hardware complexity while offering a good linearization performance.

2.3 Conclusion

The DPD design is a very important part of the communication system to provide a higher quality stream of data to the users. The DPD design presents more challenges for upcoming generations of wireless connections as better signal quality as well as higher data rates are in demand. With increasing the signal bandwidth and stricter timing requirements, the conventional DPD techniques are not able to meet the standard requirements. The DPD design trend is now moving toward the artificial intelligence-based one as AI is more flexible and can provide higher quality signals compared to conventional techniques.

3 Chapter 3: MIMO beamforming platform design and calibration

3.1 MIMO beamforming platform

The increasing need for higher internet speed and an enormous number of simultaneous users led to developing massive MIMO beamforming transmitters. The task of designing massive MIMO beamforming transmitters involves many challenges as they are composed of several RF channels in a compact area that increases the crosstalk and coupling effect. Moreover, commercial products used to design MIMO platforms have imperfections causing error and unexpected behaviour of the platform. Therefore, a systematic approach to calibrate and validate the performance of MIMO beamforming platforms is vital. Having a calibrated MIMO beamforming platform is also of great importance for proper functioning of the MIMO system. Indeed, due to the complexity of the system and the interaction and coupling between the different elements in MIMO beamforming transmitters, signal processing algorithms DPD and other algorithms are not reliable.

Generally, there are three techniques to develop MIMO beamforming platforms: using instruments, commercially available or custom design circuit boards for active phased array antenna or using commercial Software Defined Radio platforms. Designing a MIMO beamforming platform using signal generators is very expensive and impractical. In addition, it is hard to synchronize these instruments in time and frequency domain. On the other hand, using integrated options such as active array antenna may be cheaper, but it does not offer enough flexibility and re-configurability. The other approach is designing a MIMO wireless system using the commercial SDR platforms and components. Designing a MIMO beamforming DPD using commercial SDR

platforms and processing units is more challenging compared to other solutions and requires system calibration. The authors in [29] have tried to provide the design steps of an integrated MIMO platform. However, they have not used the available commercial products. There are other works trying to introduce a calibration procedure or design techniques [30][31], but none provide a systematic design and calibration procedure. The absence of a comprehensive procedure to design and calibrate MIMO beamforming platforms is the main motivation of this chapter. To address this gap, this chapter discusses a detailed procedure to design MIMO beamforming platforms using SDR solutions.

The wireless MIMO beamforming platform design is discussed in four sections: section 1 discusses an overview of the designed MIMO beamforming platform and the FPGA design, section 2 discusses the calibration steps and phased array antenna, and the last section presents the implemented MIMO beamforming platform.

Figure 3.1 A top view of the designed MIMO beamforming architecture.

3.2 5G Massive MIMO Transceiver Platform

Figure 3.1 shows an overview of the designed MIMO beamforming setup. The platform is comprised of 4 MEGA BeeCubes. MegaBeecube is an SDR platform designed by National Instruments for research purposes. Each MEGA BeeCube includes 2 AD9361 Software Defined Radio (SDR) [32] platforms and 2 Zynq 7000 series. AD9361 is a 2×2 transceiver with 12-bit resolution Digital to Analog Converter (DACs) and Analog to Digital Converter (ADCs) and it is able to support a frequency range between 70MHz to 6GHz [32]. Each AD9361 is controlled using Zynq 7000 series [33] and can be configured as a receiver or transmitter. Zynq offers both the fast-processing capabilities of the FPGA and software processing of ARMs. Using the Zynqs and AD9361 boards, all the RF paths are digitally controlled providing an ideal platform for algorithm verification.

3.2.1 FPGA implementation of the transceiver

The signal processing algorithms for RF systems are usually implemented on FPGA. Due to the high sampling frequency of base-band data, usually the signal processing algorithms are pipelined and processed sample-by-sample. This makes these designs to be delay-sensitive. Often, even one sample misalignment in processed data results in system failure. The FPGA designs for RF systems are usually implemented using System Generator- MATLAB tool [34]. This tool provides the users with pre-designed signal processing blocks with specific path delays. This feature helps when implementing the signal processing algorithms, even without a deep knowledge of FPGA architecture.

The designed MIMO beamforming FPGA platform is developed using the System Generator and it is composed of four independent FPGA designs with synchronization ports. In addition, each FPGA design contains an AD9361 control block, which is responsible to interface with the AD9361 transceiver and signal processing blocks.
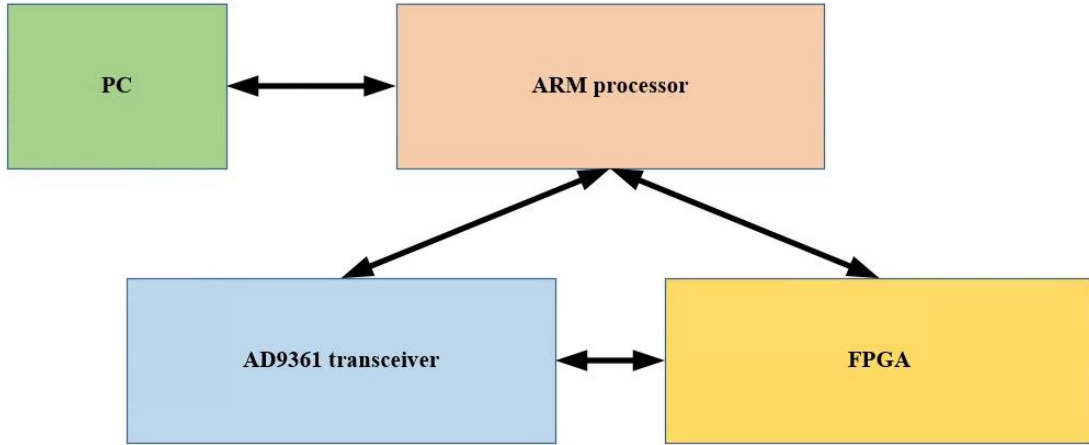
Figure 3.2 A system overview of the digital system used to develop MIMO beamforming platform.

3.2.1.1 FPGA design flow

Figure 3.2 shows an overview of the digital system interfacing AD9361 SDR, ARM, FPGA, and PC. The PC can directly send and receive information from the ARM processor in Zynq. A Linux operating system is running on the ARM processor and the user can use Putty [35] or any other software to send commands to the Linux operating system. The AD9361 is connected to the ARM processor through an SPI connection and the user is able to configure the AD9361 using Putty. The ARM processor acts as a gateway between the PC and FPGA to configure, transmit and receive data from the AD9361 SDR.

To communicate with the AD9361 SDR, a series of dedicated blocks are developed by Beecube. The control units in these blocks are directly connected to the AD9361 control registers which help the developer to control the SDR platform using the Simulink blocks. The rest of the Xilinx blocks used in the FPGA design are described in [34].

FMC112 Control: this block is a control interface to AD9361, and it is directly connected to the transceiver physical pins. Using this block RF port can by dynamically configured as input or output.

FMC112 RX: this is an interface with two channels configured as receiver. This block outputs the raw data captured by the ADC. The output format is a 12-bit two's complement integer, and it is valid every four cycles.

FMC112 TX: this is an interface with two 12-bit DAC i.e. transmitters channel. The input to this block is a 12-bit two's complement integer.

Read-Write Software Register: this is a 32-bit readable-writeable register. The software can read or write this register, but the hardware can only read it.

 GPIO: this is General Pin In-Out (GPIO) that can be configured either as an input or output and it is directly connected to the FPGA. This is a one-bit port and can provide trigger or enable-disable signal for the FPGA board.

Shared BRAM: this block creates a partition of block RAMs on FPGA, and it is accessible by both FPGA and software. The input data is 32-bit and maximum size of the BRAM is 2^{16} rows of data.

Figure 3.3 shows the block diagram of the FPGA design to interface with the AD9361 SDR. In each FPGA design 2 FMC112-TX and FMC112-RX blocks are responsible to control the transmission and reception of the signal process using AD9361 and to transfer the data to FPGA. The Synchronization port is an input port to trigger the TX channels to operate simultaneously when it is set high. As illustrated in Figure 3.4, the trigger signal is generated in one arbitrary FPGA and it is divided into four channels to be fed to the other boards. Using this configuration,
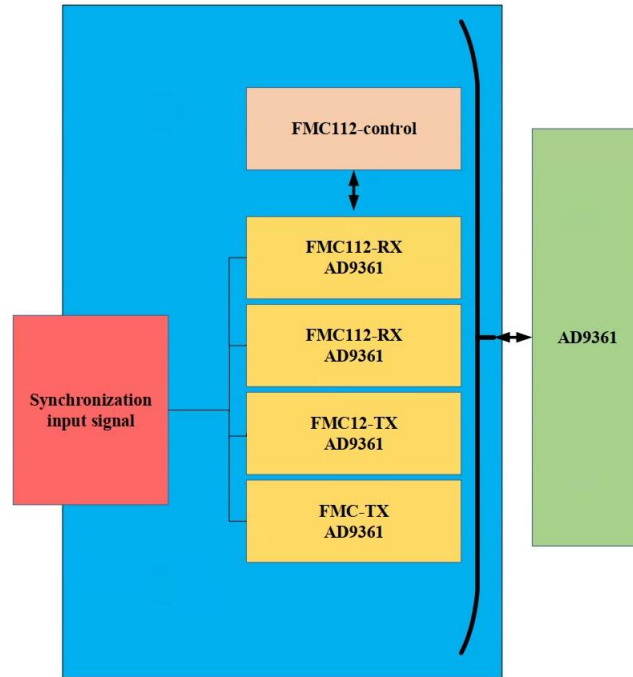
Figure 3.3 The developed FPGA design to interface the AD9361 SDR.

all the channels start the transmission at the desired instance. Without this configuration, the beamforming transmitter will not operate correctly as the channels are not time-synchronized.
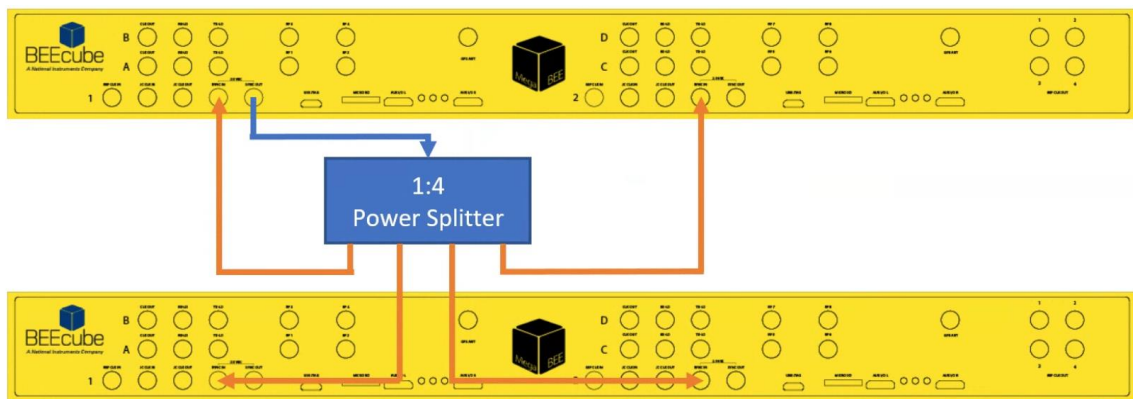
Figure 3.4 Physical interconnection between the boards for time synchronization.

3.2.2   System interconnections

For beamforming applications, two sets of phased array antenna are used to transmit signals at 2.35 GHz and 3.5 GHz. The designed platform is scalable either by adding more digital paths or connecting each digital path to a hybrid beamformer. In hybrid beamforming, each digital path is connected to a subarray and the elements in subarray are controlled using analog phase shifters. Moreover, the designed platform can operate in a multi-band beamforming scenario where each subarray transmits at a different frequency band. The final price of the designed platform using off-the-shelf products is cheaper than instruments such as vector signal generators; however, it requires calibration and impairment mitigation, which will be discussed later in this chapter.

The designed platform benefits from two observation paths: direct connection from each PA to receiver using SMA cables and a dipole far-field receiver. The far-field receiver is mainly used for DPD identification and beamforming analysis. The baseband signal is loaded from FPGA, and then pre-calculated beamforming weights are multiplied to the signal in the transmission chain. Next, the signal is up-converted, amplified, and transmitted through the phased array antenna. On the other end, the dipole receiver antenna captures the far-field signal; then it is down-converted and passed to the FPGA and PC for further analysis. The transmitted and received signal are used to extract the DPD model used to enhance the signal quality.
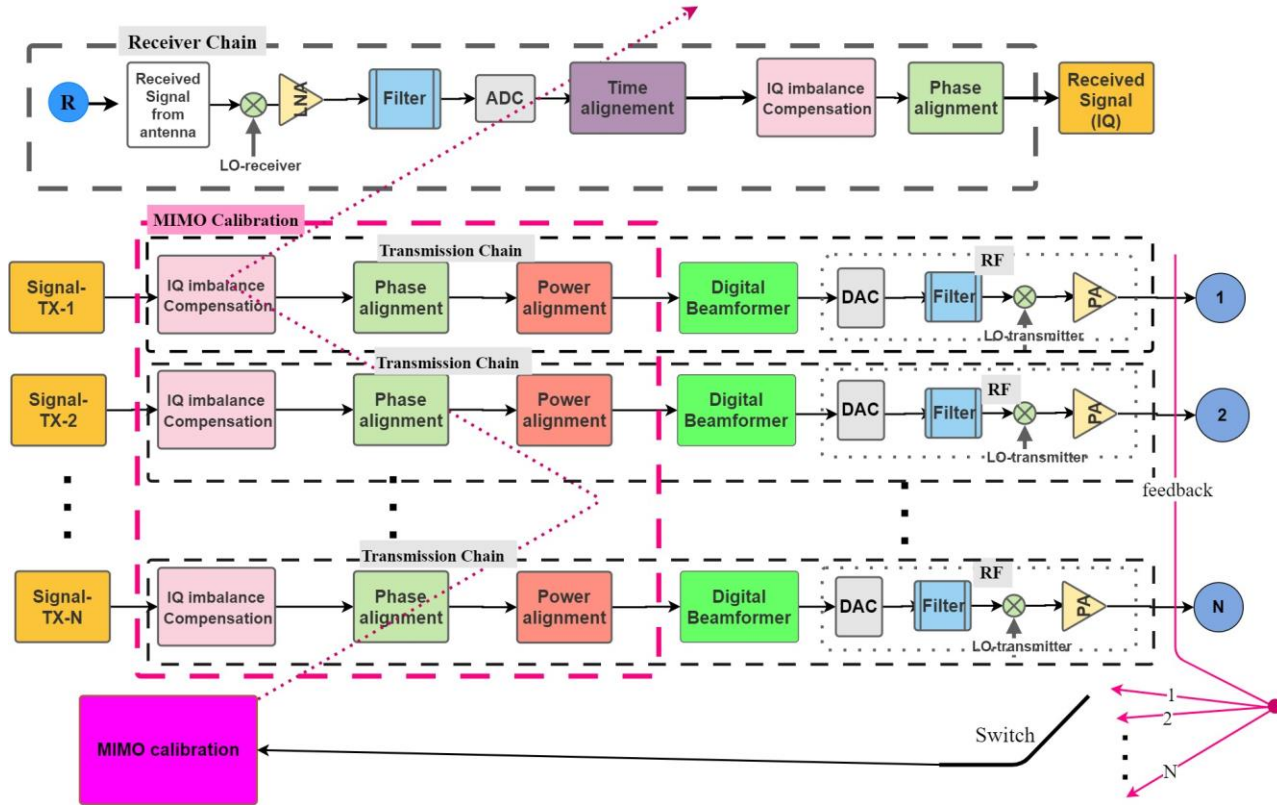
Figure 3.5 Detailed view of the transmission and receiver chain of the designed MIMO beamforming architecture.

Figure 3.5 shows the different components of the transmission and receiver chain in detail. The original signal passes through a complex conjugate filter to remove the IQ imbalances and DC offset. Then, the signal is phase corrected to compensate for phase difference between the RF chains. There is a power control unit to set the output power to the desired value. Afterward, digital weights are introduced to the signal for beamforming purposes. Next, the signal is converted to analog, up-converted, amplified and transmitted through the phased array antenna. The receiver path is composed of a Low Noise Amplifier (LNA), Automatic Gain Controller (AGC), down converter and an ADC to send the signal to the FPGA. In FPGA, the signal is phased and time-aligned with a reference signal, filtered and passed to PC for further analysis.

3.3   Platform Calibration

For the MIMO beamforming platform to operate as expected, it needs to be calibrated. Figure 3.6. briefly summarizes the required calibration steps for the designed MIMO beamforming platform. The first step is to develop the FPGA design; then, the PA output power should be set to the desired value. The commercial IQ modulators contain imperfections and as a result, they introduce IQ imbalance and DC offset. For a higher signal quality, the IQ imbalances of the RF boards are removed using a complex conjugate filter. For a beamforming transmitter to operate properly, the RF channels must be frequency coherent. Since the RF boards use different Phase Locked Loop (PLL) to generate the RF frequencies, the phase of the generated frequencies are random. Therefore, there will be a constant phase offset between the output RF signals. To have identical RF outputs or, in another word, to have two coherent RF outputs from the transceivers, this phase offset should be compensated. The other aspect of frequency coherence is the frequency stability. For the two RF signals to be coherent, the center frequencies should be stable and identical during the time. There are two ways to have beamforming transmitters coherent: using an external Local Oscillator (LO) or using a common reference clock and feeding to the RF boards to generate their own LOs. Finally, to validate the quality of the calibration and frequency coherence of the transmitter, the phase offset compensation and cancellation test is carried out. In addition to validating the calibration, the cancellation test achieves phase coherence. You need to tune the phase difference between the paths to have a good cancellation result. When using a
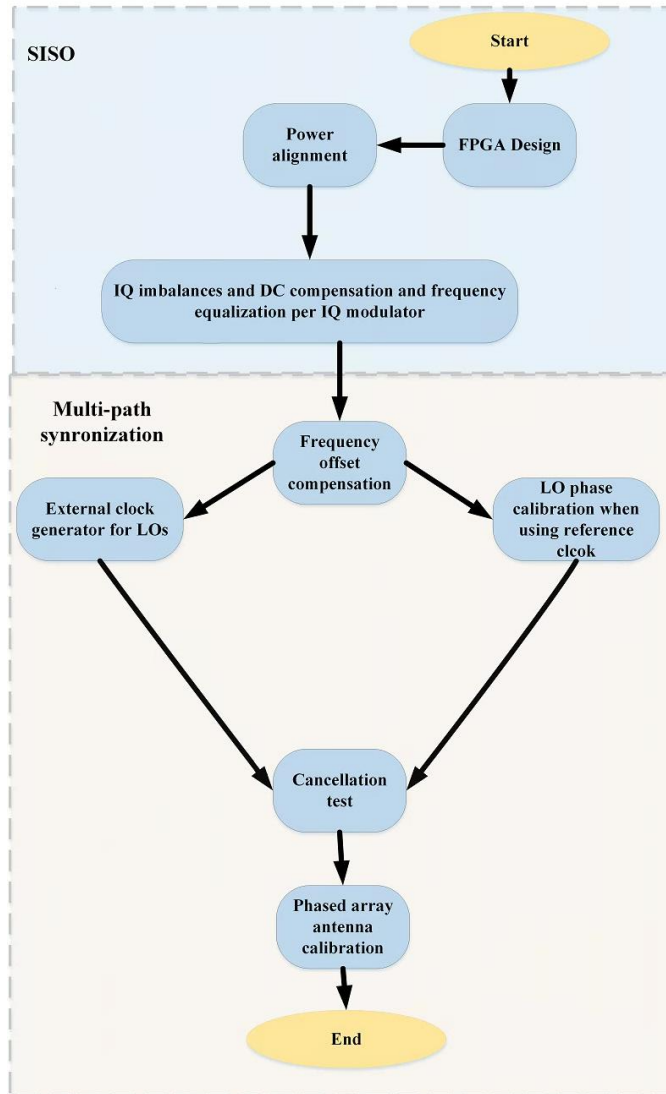
Figure 3.6 Flowchart summarizing the required calibration steps for MIMO wireless systems with active antenna array.

common reference clock, frequency coherence only guarantees that LOs have the same frequency, but not necessarily the same phase. In the following sections, each of the MIMO beamforming design and calibration steps are discussed.

3.3.1   IQ imbalances and DC offset

As massive MIMO systems are designed using a commercial RF up-converter and down-converter, they suffer from manufacturing imperfections causing IQ imbalance and DC offset. IQ imbalance can degrade the signal quality and beamforming accuracy [36]. The IQ imbalance is expressed as follows:

$$x_{out-mixer} = £.x + ¥.x^*$$
(3.1)

where £ and ¥ are scaler numbers function of the amount of IQ imbalance, x is the complex envelope of the baseband I and Q input signals to the IQ modulator, x^* represents the conjugate of x and $x_{mixer-out}$ is the complex envelope of the mixer output. [37].

A first-order complex conjugate filter with M memory taps is used to remove the IQ imbalances. The memory depth of the complex conjugate filter is determined based on the behaviour of the system. In beamforming setup, there are two ways to extract the complex conjugate filter coefficients, either using direct measurement or the subarray technique. Figure 3.7 (a) shows the architecture of IQ imbalance direct measurement technique. In the direct measurement, the RF output of each channel is directly connected to the receiver, and the received signal is used to extract a complex conjugate filter to remove the imbalances in that chain. This is the most accurate way; however, it is only applicable to fully digital platforms. The other technique is receiving the far-field signal and extracting a complex conjugate filter per subarray. Figure 3.7. (b) shows the architecture used to identify one IQ imbalance filter per subarray. In this technique, the far-field signal, which is the sum of all PAs in one direction, is received and used to identify the IQ imbalance filter. Table 3-1 reports the quality of a signal before and after applying the complex conjugate filter for different techniques operating in linear

region. The results show that using one complex conjugate filter per subarray can offer a relatively close performance to the case where each RF channel is *IQ* imbalanced using an independent filter.

Table 3-1 Effect of the complex conjugate filter on the IQ imbalances compensation.

Filter memory (M)	NMSE(dB)
No filter	-26.1
Filter per channel M=3	-30.2
Filter per channel M=5	-28.3
One per subarray M=3	-29.1
One per subarray M=5	-28.1
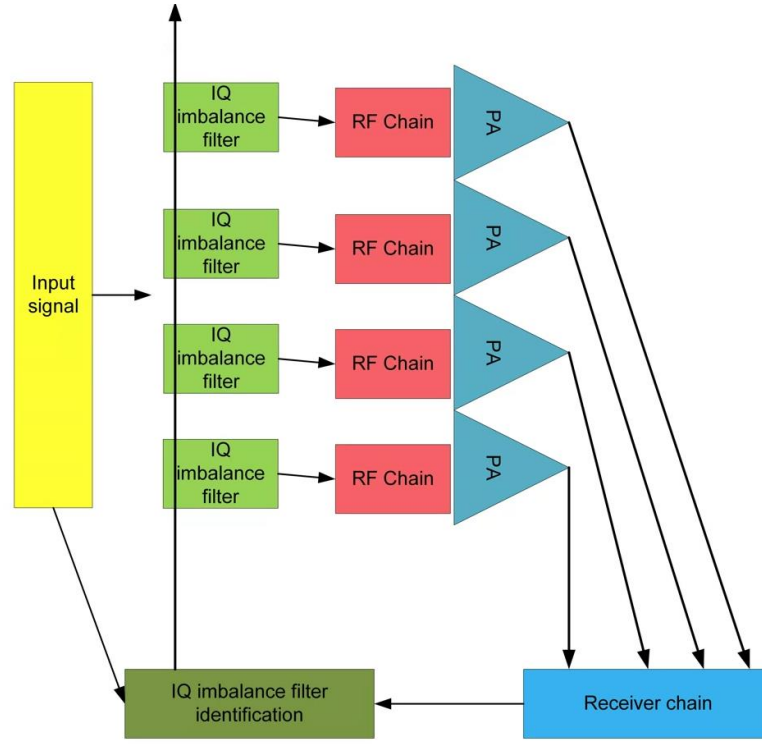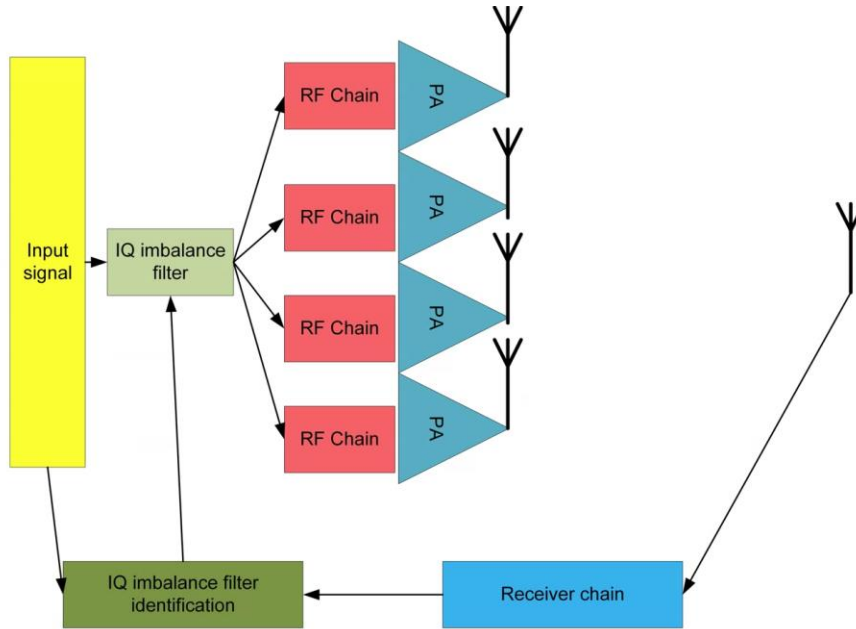
Figure 3.7 (a) IQ imbalance filter identification using direct calculation one filter per RF channel, (b)IQ imbalance filter identification using the far-field signal one filter per subarray.

3.3.2   Multi path synchronization

3.3.2.1   Frequency offset

Commercial transceivers suffer from imperfections and frequency offset between the desired and actual carrier frequency. To mitigate this frequency offset, a 50 kHz single tone signal is transmitted at the desired center frequency and the difference between the actual output frequency and the desired one is used to tune the center frequency to the desired one.

3.3.2.2   Local oscillator phase calibration

The Local Oscillator (LO) directly affects the signal quality as it converts the signal from base-band to radio frequency or vice versa. There are four aspects about the LO that need to be considered: the quality of the LO signal, the frequency stability, frequency offset, and the phase coherence of two LOs. Frequency stability means that the frequency does not drift over time. Phase coherence means that LOs have the same phase. As shown in Figure 3.8, there are two ways to design a phase-coherent system. Using an external LO generator is the most straightforward way; however, it makes the system bulkier and more expensive. Moreover, by increasing the number of boards, higher output power from the LO circuit is needed which increases the LO phase noise and reduces the signal quality. The other way to design a phase-coherent transmitter is using a common reference clock and feeding it to the PLL inside the RF boards to generate their own LOs. This technique is more practical for industrial applications; however, it requires more calibrations. As each transceiver has an internal Phased Locked Loop (PLL) to generate the LO signal, each LO signal has a different phase. In order to have two coherent RF transmitters, the phase offset between
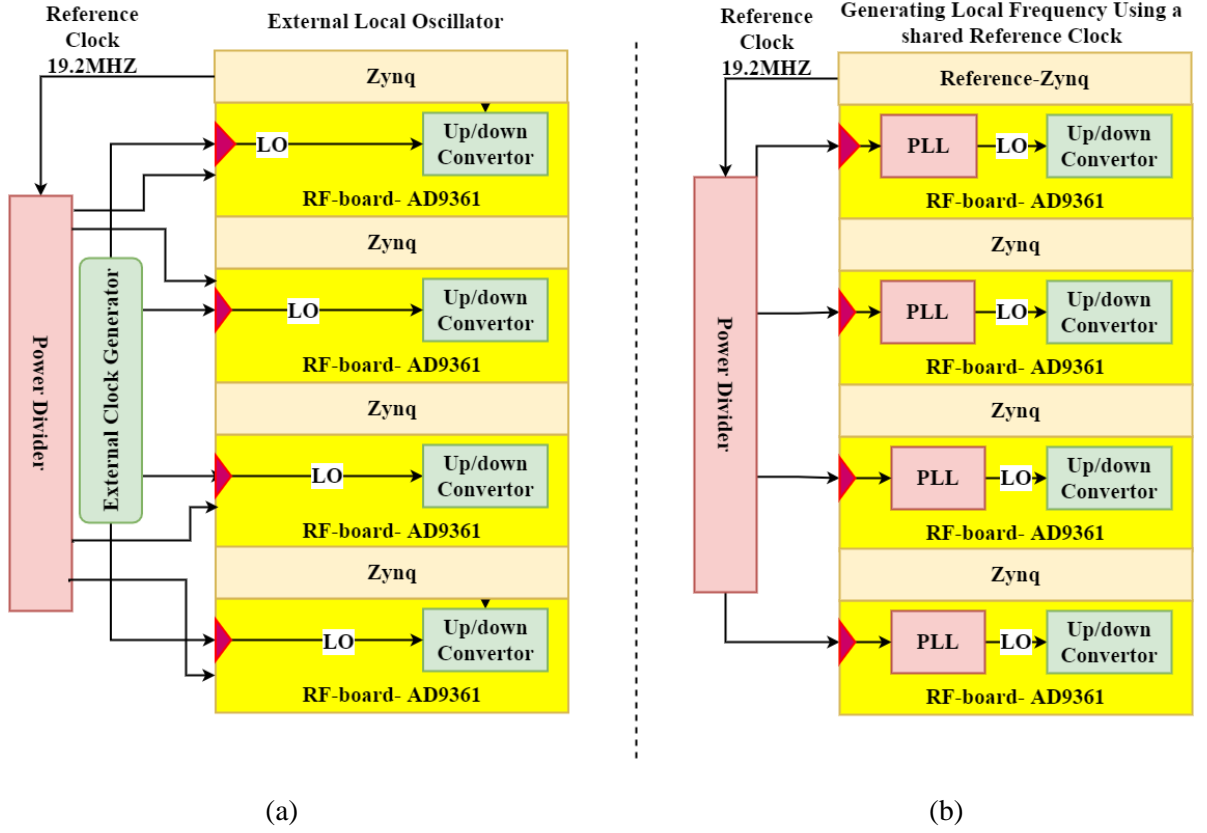
Figure 3.8 Typical techniques for Clock distribution in RF boards a): external LO, b): external reference clock.

them should be estimated and compensated. The phase difference $\Delta\widetilde{\emptyset}$ between two RF transmitters, x_1 and x_2 is calculated as follows:

$$x_1 = x_i^1.\cos(2\pi f_{LO}t + \epsilon_1) + x_q^1.\sin(2\pi f_{LO}t + \epsilon_1) \qquad (3.2)$$

$$x_2 = x_i^2.\cos(2\pi f_{LO}t + \epsilon_2) \ldots + x_q^2.\sin(2\pi f_{LO}t + \epsilon_2) \qquad (3.3)$$

$$\Delta\widetilde{\emptyset} = \text{mean}\left(\text{phase}\left(\frac{x_2}{x_1}\right)\right) \qquad (3.4)$$

Where x_i and x_q are the in phase and quadrature part of the original signal x, ϵ_1 and ϵ_2 are the phase offset of LO of the transmitter 1 and 2.

There are two ways to estimate the phase difference between RF channels; direct measurement and over-the-air measurements. In the direct measurement technique, the output of each RF channel is directly connected to the receiver through an SMA cable. Then, the phase offset between the original signal and the received one is estimated using equation (3.4) and loaded into the transmission path. Direct measurement is the most accurate way, but not the only way. If the RF channels are not directly accessible, over the air measurement can be used. In this technique, only one radiating element is turned on at a time, and the far-field received signal is acquired. It is important to note here, that due to different physical distances between each element and the far-field received signal, a phase offset is introduced to the received signal. The phase offset for the element in m^{th} row and n^{th} column relative to the reference element shown in Figure 3.9 is estimated as follows [38]:

$$W_{m,n} = \sum_{m=1}^{M} \sum_{n=1}^{N} e^{+j\cdot\left(2\pi\cdot\frac{d_x}{\lambda}\cdot(m-1)\cdot\sin(\theta)\cos(\phi)\right)+j\cdot(2\pi\cdot\frac{d_y}{\lambda}\cdot(n-1)\cdot\sin(\theta)\sin(\phi))} \tag{3.5}$$

The λ is the signal wavelength, ϕ is the elevation angle and θ is the azimuth angle. d_x, d_y are the distances between the adjacent elements in x-axis, y-axis, respectively.
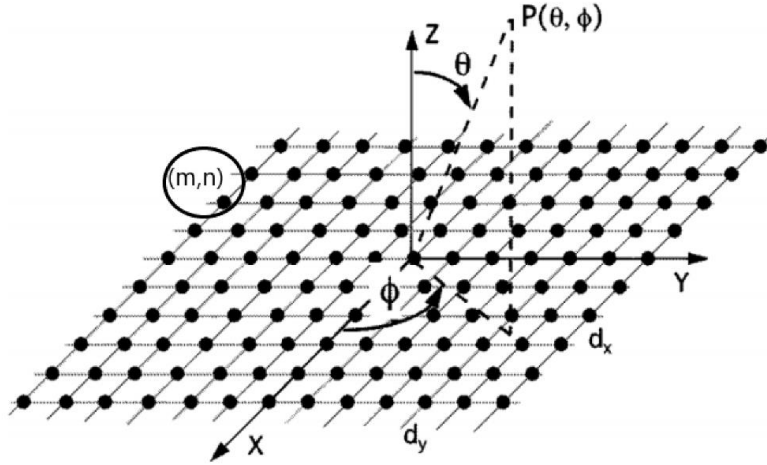
Figure 3.9 Phased array antenna with d_x, d_y distance between the element is x and y direction, respectively.

First, the phase offset estimated using equation (3.5) should be excluded from the received signal and then the phase offset between the RF chain and the reference channel (which is the receiver) is calculated. It is worth noting that all the RF channels are phase-aligned relative to the receiver chain. The estimated phase offset is then loaded to the phase corrector register in the FPGA design shown in Figure 3.5.

3.3.2.3 Phase offset compensation and Cancellation test

In order to verify the coherence of the beamforming transmitter, it should be tested using a cancellation test. All the RF channels in a subarray should be aligned in power, and coherent in frequency and in time. The time coherence means that all the RF channels transmit the same data at the same moment. As mentioned earlier, a time synchronization port has been developed and
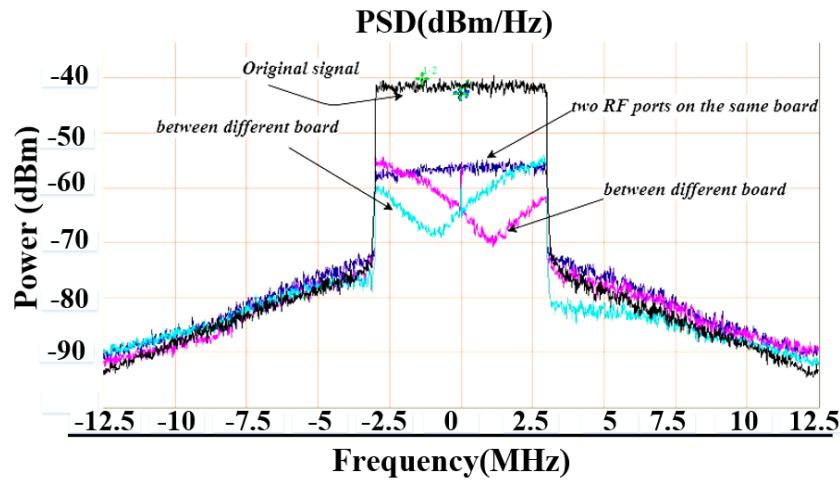
Figure 3.10 Cancellation test results acquired from two RF port mounted on the same RF board and two RF ports mounted on different RF boards.

used to trigger the transmitter simultaneously. The following instructions describe the cancellation test:

1. The PAs are tuned to output the same power

2. One signal is shifted by 180° and the other is transmitted without any change

3. Two signals are summed in a combiner

4. The output spectrum is displayed on a spectrum analyzer

Figure 3.10 shows the output spectrum of the combined signal for two scenarios. If the RF front-ends are on the same board, they use the same phase locked loop (PLL). In this case, phase and frequency are coherent between the RF front-ends and the output spectrum is flat. Ideally, the combined signal should be close to the noise floor but due to the controllability of the PAs and LOs, an average result of -15 dB cancellation is acceptable.
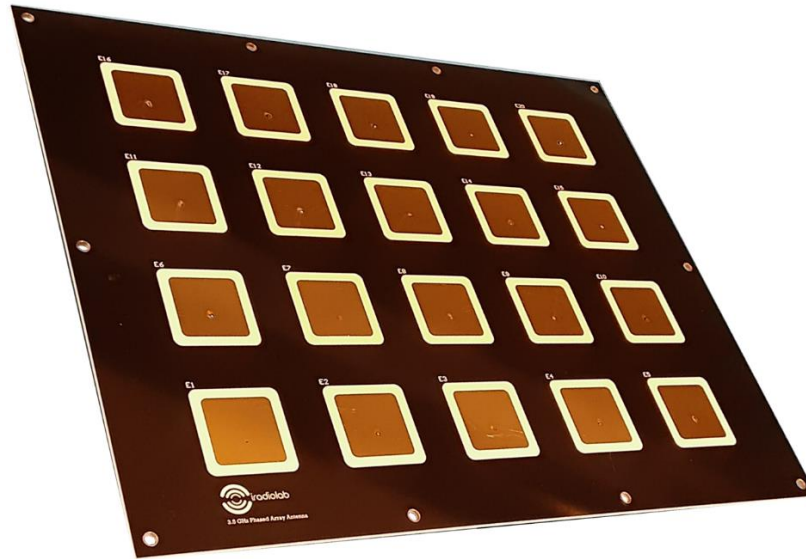
Figure 3.11 Phased array antenna design to operate at 3.5 GHz center frequency.

3.4 Phased array antenna

The Beamforming technique is used to steer the transmitted signal in the direction of the receiver. Using this technique, the interference is reduced, the link security is improved, and spectral efficiency is enhanced. In beamforming, the data is transmitted using a phased array antenna. Figure 3.11 shows a phased array antenna used to transmit data at 3.5 GHz. A phased array antenna can be divided into groups of elements called subarrays. Each subarray can transmit the data independently from the other subarrays. There are three common beamforming techniques: analog beamforming, hybrid beamforming, and digital beamforming. Figure 3.11 shows the architecture of the analog beamforming. In analog beamforming, an analog phase shifter is used to introduce a phase offset to each RF path. The other technique is called hybrid beamforming. Figure 3.12 illustrates the architecture of the hybrid beamforming. In hybrid
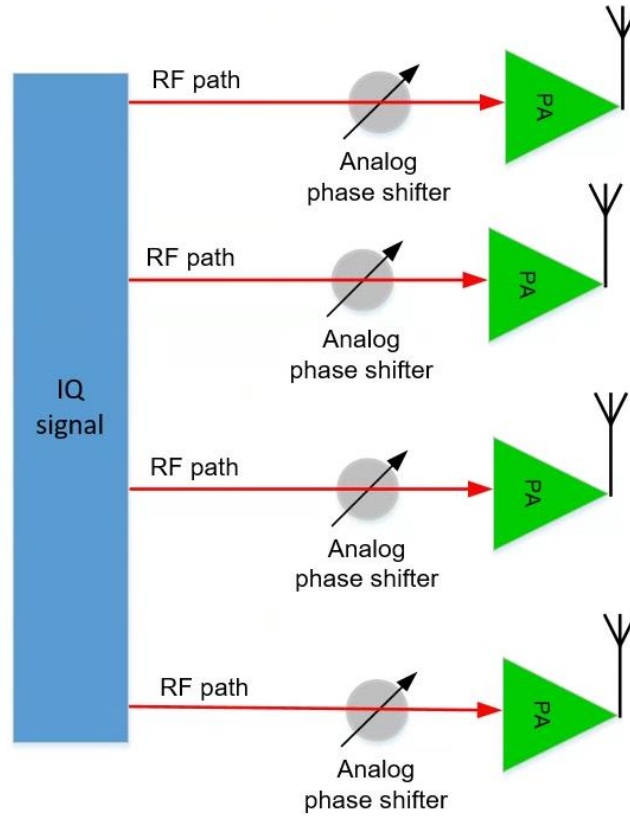
Figure 3.12 Architecture of the analog beamforming transmitter.

beamforming, the phase offset to each RF path is introduced using a digitally controlled analog phase shifter. The beamforming can also be done using fully digital transmitters too. As shown in Figure 3.13, in fully digital beamforming, each RF path is connected to a digital path and the appropriate phase offset is introduced on the baseband signal. Currently, hybrid beamforming [39] and digital beamforming have drawn attention from both academy and industry. However, the digital beamforming is more accurate and suitable for deployment as it provides more controllability. This chapter focuses on the digital beamforming.
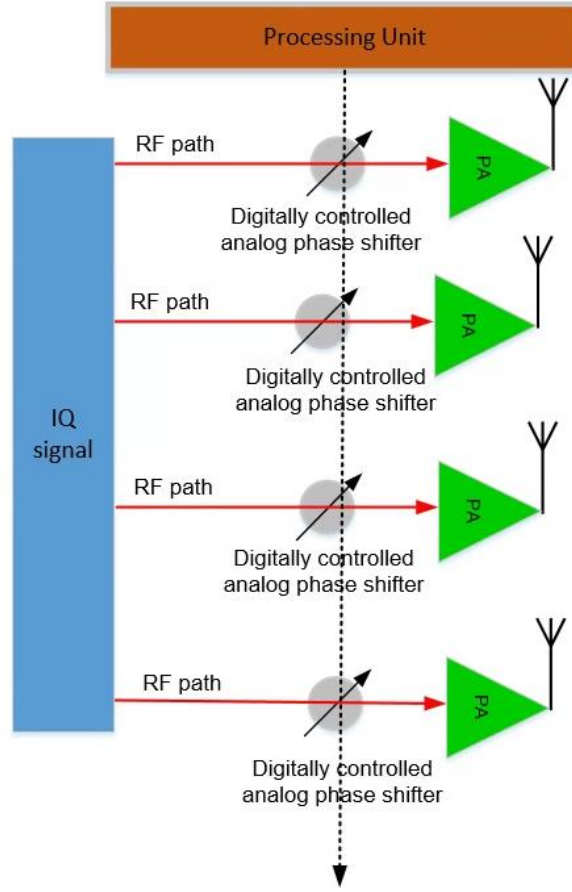
Figure 3.13 Architecture of the hybrid beamforming transmitter.

3.4.1 Beamforming

The basic concept behind the operation of beamforming is that the distance between each element and the receiver is different and this causes a phase offset between the received signal from each radiating element. The received signal from the element in the mth row and n^{th} column is calculated as follows [38]:

$$\tilde{x}_{m,n}(t)$$

$$= x(t) . \sum_{m=1}^{M} \sum_{n=1}^{N} e^{j\left(2\pi.\frac{dx}{\lambda}.(m-1).\sin(\theta)\cos(\phi)+(m-1).\theta\right)} \times e^{j\left(2\pi.\frac{dy}{\lambda}.(n-1).\sin(\theta)\sin(\phi)+(n-1).\psi\right)} \quad (3.6)$$

with
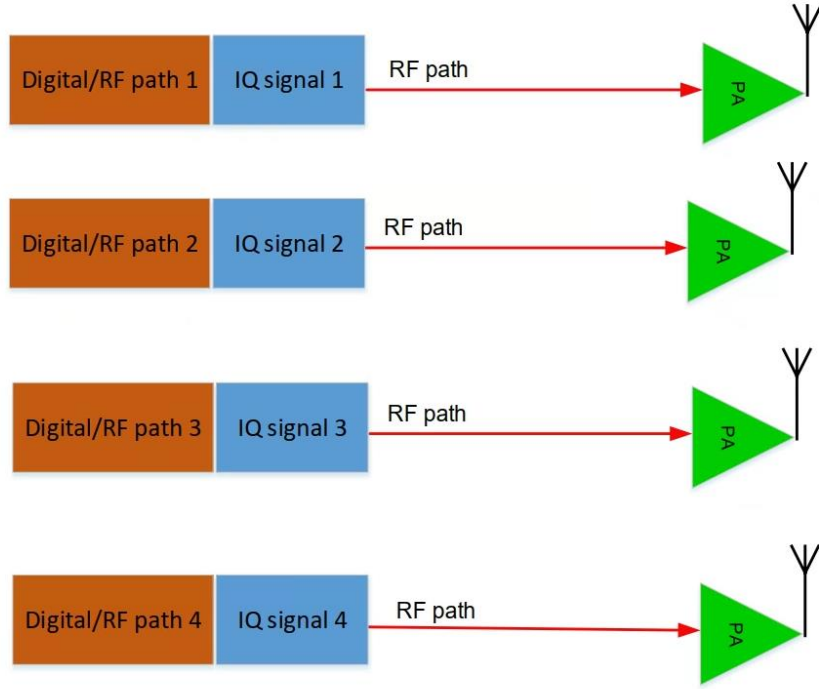
Figure 3.14 Architecture of the digital beamforming transmitter.

$$-2\pi . \frac{dx}{\lambda} . \sin(\theta) \cos(\phi) = \theta \tag{3.7}$$

and

$$-2\pi . \frac{dx}{\lambda} . \sin(\theta) \sin(\phi) = \psi \tag{3.8}$$

The λ is the signal wavelength, ϕ is the elevation angle and θ is the azimuth angle. d_x, d_y are the distances between the adjacent element in x-axis, y-axis, respectively.

Equation (3.7) estimates the phase offset between two signal paths. The first path is from the reference antenna element of the antenna array, which is the element in the first row and first column to the receiving antenna, to the receiving antenna. The second is from another antenna element to the receiving antenna. To steer the beam in a given direction, the received signals from

all the transmitting antenna array elements should be in-phase. To do so, the transmitted signal in the second path is shifted by the opposite of the phase difference obtained by equation (3.7). There are different beamforming algorithms and techniques [40] [41] to increase the accuracy of the beam steering, which are not the focus of this thesis.

3.4.2   Phased array antenna specifications

The developed MIMO beamforming setup transmits data using two 5×4-element phased array antenna operating at 3.5 GHz and 2.35 GHz center frequency using two different antenna arrays. The 3.5 GHz phased array antenna has a gain of 17 dB and beamwidth of 27°. The 2.35 GHz phased array antenna gain is equal to 12 dB and beamwidth is 34°. Figure 3.15 shows the measured radiation pattern of the 2.35 GHz phased array antenna when it is steered toward the 0° and -30°. The phased array antenna should be tested and calibrated for any phase or gain offset between the different paths of the transmitted signal. The RF signals in the phased array antenna used in the developed MIMO beamforming platform are directly connected to the back of the radiating element to eliminate any phase or gain variation of the transmitted signal. However, the phase and gain offset can be generated from different sources such as cables, connectors, and RF circuitry. Therefore, phase and gain calibration are important to increase the beamforming quality.

3.5   Designed MIMO beamforming platform

Figure 3.16 (a) shows the architecture of the designed MIMO platform. A 19.2 MHz reference clock generate from one RF board is power divided and fed to other boards. The RF
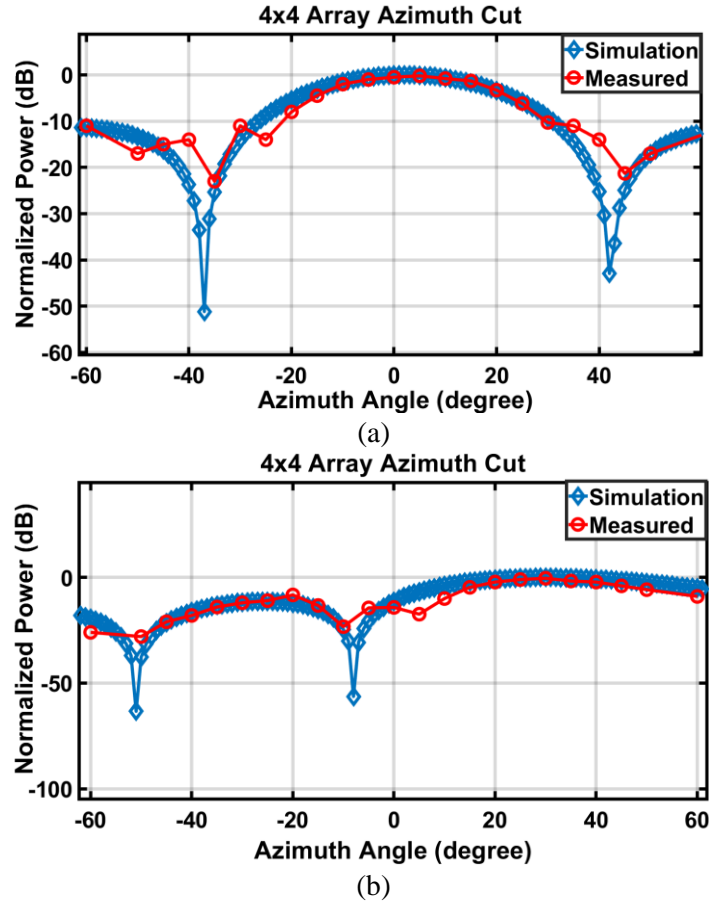
Figure 3.15 Azimuth cut of 4x4 phased array antenna for both simulation results and measured value at; a) 0°,b) 30°.

boards use this reference clock to generate their own LO using the built-in PLL. A time trigger

signal is also generated in one board and fed to the other three to synchronize them.

The phased array antenna is connected to the AD9361 boards. To reduce the reflections

and acquire repeatable results, the far-field receiver is fixed in an absorber cage and the transmitter

(a)

(b)
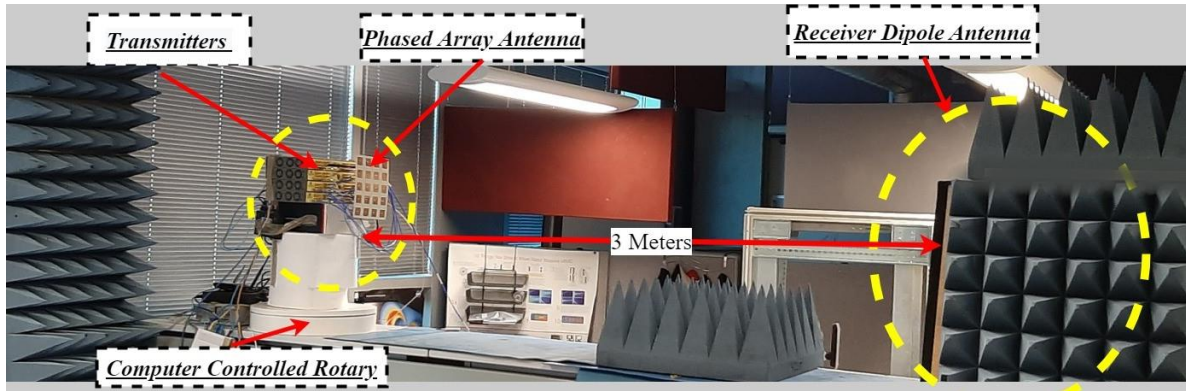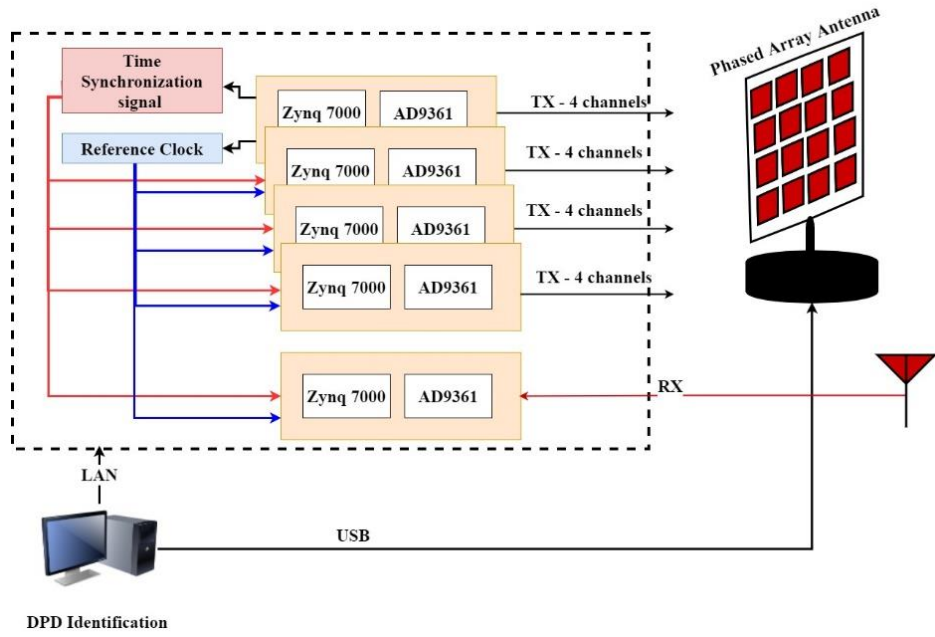
Figure 3.16 (a): Schematic of the 16-channel MIMO digital beamforming with over the air signal acquisition. (b): Implemented version of the MIMO digital beamforming platform.

is rotated in the inverse direction of the beamforming using a computer-controlled rotary. The receiver is set 3 meters away to capture the far-field signal. Figure 3.16 (b) shows the implemented setup in the laboratory.

The system is calibrated based on the instructions given in Section 3.3. In the first attempt to analyze the performance of the signal, the signal is transmitted toward 25°angle and the far-field

signal is received with the horn antenna. The received signal is then used to identify a DPD with 3 memory taps and a nonlinearity order equal to 12. The power spectrum of the calibrated and noncalibrated transmitted signal are shown in Figure 3.17. The results show that the calibration increases the dynamic range and quality of the signal. The DPD extracted using a calibrated system achieves a linearization performance with -48 dBc Adjacent Channel Power Ratio (ACPR) while only -45 dBc ACPR is obtained with a noncalibrated system. One way to observe the impact of calibration on the signal quality is by plotting the spectrum of the error signal, which is the difference between the original signal and the output of the transmitter. The results show that the calibrated system offers better performance. The Normalized Mean Square Error (NMSE) of the difference between the original signal and the received one for calibrated and uncalibrated systems are -34.3 dB and -26.7. The NMSE of the difference between the transmitted signal and the received one is calculated using equation (2.2) presented in chapter 2.
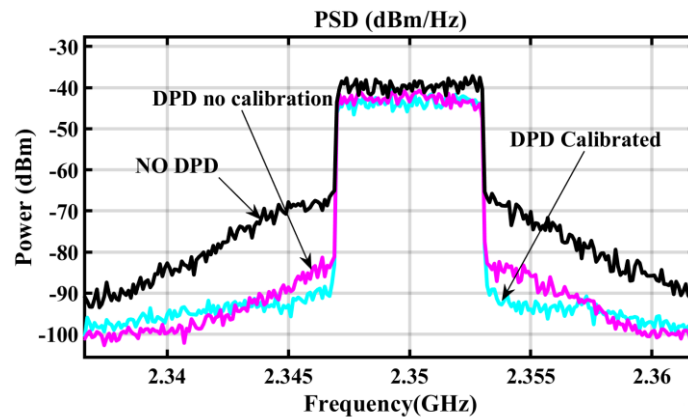
Figure 3.17 Spectral power of the received signal with and without calibration received using power spectral analyzer.

3.6   Conclusion

The need for a detailed procedure to design MIMO beamforming platform motivated this chapter. This chapter proposed a detailed calibration and design procedure to develop MIMO beamforming platforms using off-the-shelf commercial products. The FPGA design and blocks to interface the SDR platform, and PC are discussed. Two techniques to mitigate the IQ imbalances and the DC offset were introduced. The system interconnections, including the LO connections and synchronization signals are developed. Finally, the cancellation test to observe the system coherence in time and frequency is carried out. The measurement results using the developed MIMO beamforming platform shows that the platform can provide an acceptable performance in terms of NSME and ACPR of the transmitter signal. This developed and calibrated MIMO platform will be used in the remainder of the thesis to test and validate signal processing algorithms to improve the linearization performance in MIMO and beamforming transmitters.

4 Chapter 4: Angle inclusive beamforming DPD for phased array antenna

4.1   Introduction

Massive MIMO beamforming platforms suffer from nonlinear distortion that varies as a function of the transmission direction in addition to the severe crosstalk between the radiating elements. Moreover, the bulkiness of the base stations is an issue that needs to be addressed. To reduce the number of digital chains in beamforming transmitters, hybrid beamforming is used. Using this technique, only one digital chain is used per subarray. There are two remaining bottlenecks, the hardware complexity and the online identification of the DPD coefficients. To address this issue, this chapter proposes a novel Angle-Inclusive Digital Pre-Distortion (AI-DPD) that models the beamforming transmitter and provides uninterrupted linearization across the steering range. Moreover, it reduces the hardware complexity by eliminating the need for an independent observation path for each subarray. The proposed AI-DPD is implementable using Artificial Intelligence (AI) to model the behaviour of the beamforming transmitter across the steering range. The main novelty of the AI-DPD is the ability to include and capture the effect of steering the beam angle on the transmitter's nonlinear behaviour.

AI has shown to be a very powerful and handy technique when it comes to problems where scientists cannot devise a solid mathematical representation to describe the problem or when the mathematical representation is too complex for practical applications. Despite the common belief that AI algorithms increase the system complexity, in this chapter we will prove that AI algorithms can reduce the complexity in massive MIMO, if it is designed for the right application. With the increasing applications of AI in the modern world, many vendors are providing open-source code and libraries as well as hardware to carry out the AI computations faster and more efficiently [42]. There are different types of AI models and studying them is out of the scope of this thesis. This
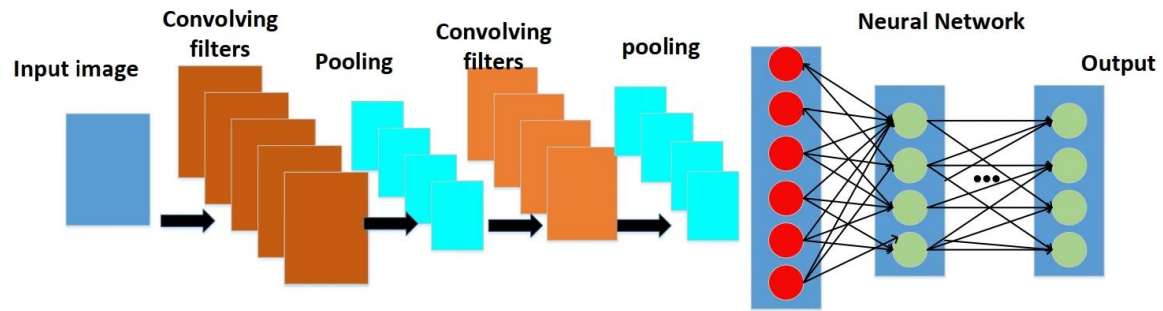
Figure 4.1 Top view of a typical convolutional neural network architecture.

chapter focuses on designing the AI-DPD using Convolutional Neural Network (CNN). CNN has proven to be able to provide strong modeling capabilities while reducing the complexity compared to other artificial intelligence models. It also reduces the model identification complexity by reusing the parameters of the convolutional layer in different operating conditions. To review the CNN model, the following sections discusses the CNN architecture and its operation concepts in details.

4.2   Convolutional Neural Network

The primary application of the CNN is to categorize the images that a machine receives. CNN reduces the gap between human and machines by enabling the computer to distinguish different objects. The CNN is famous for its capability to extract image features and categorize the image features which is widely used in autonomous car and other image processing applications. However, its application is not limited to image processing, and it can be used in variety of applications.

Figure 4.1 depicts the general architecture of a typical CNN model. The CNN layers are similar to deep neural networks. The main differences are convolutional layer and pooling layer. Convolutional layer is composed of a series of convolving filters responsible for extracting image features. The pooling layer is responsible for down sampling the data and reducing the data size. The output of the last pooling layer is passed to a fully connected neural network to generate the output data. The output data can be used either for classification or generating an output number.

CNN is trained using input-output pairs of the system to train the filter coefficients, biases and Neural Network (NN) weights. One of the main advantages of the CNN over the other techniques, is the possibility of using the already trained convolutional layer for re-training the CNN network which is called transfer technique [43]. The other important feature of the CNN is called shared parameter in which the output of the feature maps (output of convolutional filters) use the same weight to feed the data to the fully connected neural network layer [44].

4.2.1   Input image

The input image contains the input data information in 2-D format as it is illustrated in Figure 4.2. The input image could be an image captured by a camera or another information which are organized into a 2-D array.

4.2.2   Convolutional layer

Figure 4.2 illustrates the basic concept operation of the convolutional layer. This layer is responsible for extracting the image features and passing them to the pooling layer. The convolutional layer is composed of several filters called feature extractors. In CNN convolutional layer, each filter slides over the input image, performs a dot operation on each input pixel, sums the results, and projects it to another pixel in the next layer. Figure 4.3 presents an example

showing the operating concept of the convolutional layer. An example of the input volume is shown in Figure 4.3 as an image of size 5×4. Three convolutional filters with size 3×3 are used in this example as shown in the figure. Each filter slides over the input volume and in each position, it generates one element of the output volume. For example, when the filter W0 is in the position
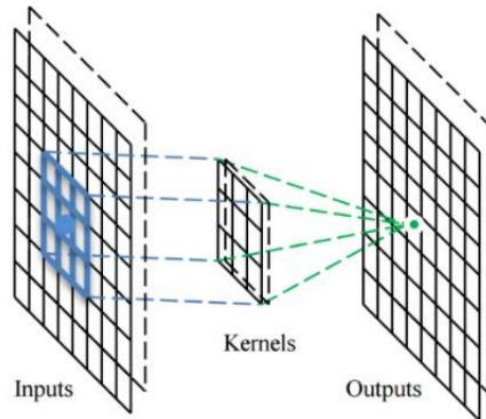
Figure 4.2 The operation concept of the convolutional layer [47].
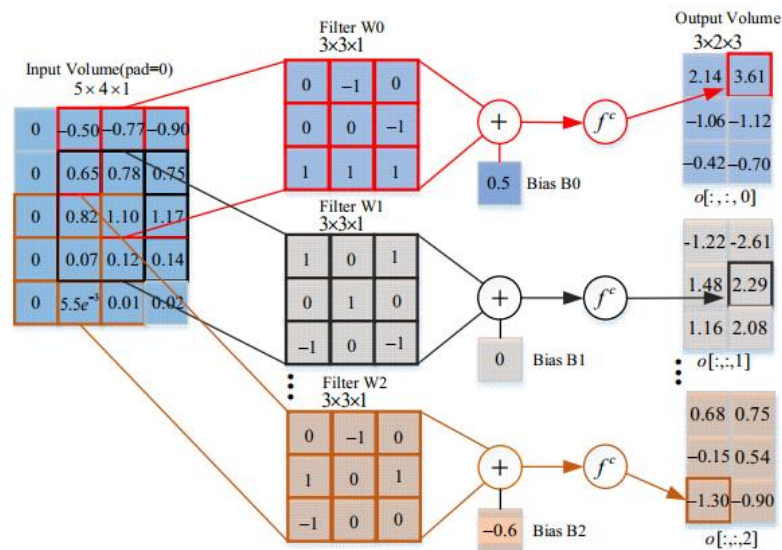
Figure 4.3 An example showing the convolutional filter operation [11].

shown in the figure, the output consists of the activation function f^c applied to the sum of the result of the convolution added to the bias as follows:

f^c ((-0.5) · 0 + (-0.77) · (-1) + (-0.9) · 0 + 0.65 · 0 + 0.78 · 0 + 0.75 · (-1) + 0.82 · 1 + 1.1 · 1 + 1.17 · 1 + 0.5) = 3.61

4.2.3 Activation function

As it is shown in Figure 4.3 the output of each filtering iteration is passed through an activation function to generate the output pixels. Activation function is a nonlinear function that generates the data for the next layer. There are different types of activation functions. The most common activation functions as given below:

$$\text{Relu: max } (x, 0) \tag{4.1}$$

$$\text{LeakyRelu : max } (\alpha x, x) \tag{4.2}$$

$$\text{Sigmoid: } \frac{1}{1+e^{-x}}. \tag{4.3}$$

$$\text{ELU:} \begin{cases} x \; ; \; x > 0 \\ \beta(e^x - 1) \end{cases} \tag{4.4}$$

Relu activation function is among the most used one as it reduces the complexity and also doesn't trigger all the neurons at the same time. This reduces the hardware and computational complexity compared to the other functions.

4.2.4 Pooling Layer

Pooling layer is responsible for reducing the data size for the downstream layers and removing the unimportant features in the feature map. The pooling layer can be either max-pooling or average-pooling function. In max-pooling techniques, the volume is divided into portions and

the maximum value of pixels in that portion is projected to the output volume. In the same manner, the average-pooling kernel maps the average value of each portion into the output volume. However, since the input image and feature maps in the AI-DPD are small, the pooling layer is omitted in AI-DPD.

4.2.5 Fully Connected Neural Network layer

Fully connected neural network is a straightforward way to convert the extracted features into a physical or classified interpretation. The output of the convolutional layer or pooling layer is flattened into a vector and passed to the fully connected layer to generate the output data.

4.3 Angle Inclusive Digital Pre-distortion

Beamforming transmitters suffer from direction dependency of the distortions. This requires DPD re-identification as the beam is steered toward the different directions. However, the DPD re-identification requires both a dedicated observation path to monitor the nonlinearity and a strong processing unit to identify the DPD coefficient fast enough. The online updating of the DPD coefficients is time consuming and might prevent a transmitter to meet timing requirements. Moreover, during the DPD re-identification, the transmitter might violate the Adjacent Channel Power Ratio (ACPR) regulations and blocks the other receivers. To both provide an uninterrupted linearization-eliminates the need for online readaptation of the DPD coefficients-and eliminates the need for a dedicated observation path, the AI-DPD is introduced in this chapter. The proposed
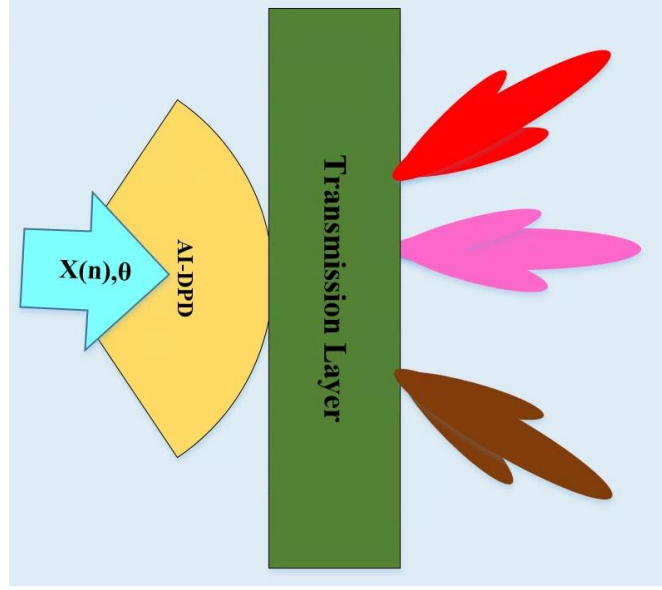
Figure 4.4 Block Diagram of a beamforming transmitter using the proposed AI-DPD.

AI-DPD models the beamforming transmitter considering the directional behaviour of the beamforming transmitter. Figure 4.4 shows the basic operation concept of the AI-DPD. It takes the steering angle as well as the modulated signal as the inputs, and pre-distorts the signal accordingly. One way to include other parameters other than the modulated signal into the DPD model is using AI, as the inputs to the AI model can be easily augmented to encompass more parameters. The proposed AI-DPD is composed of the basic layers described in section II. However, the input image and other layers are customized for the AI-DPD application. The following section discusses each AI-DPD layer in details.

4.3.1   Input layer

The AI-DPD input layer is an input image that contains the transmitter information. The input information to the model should be provided and organized in a way to both increase the modeling accuracy and reduce the model complexity. Equation (4.5) expresses the input image to

the AI-DPD model. The first input is $I(n-m)$ and $Q(n-m)$, which are the in-phase and quadrature part of the input signal. The $|x(n-m)|^q$ is the qth order of mth delayed version of the modulated input signal complex envelope.

$$Image_i = \begin{bmatrix} I(n) & I(n-1) & I(n-2) & ... & I(n-m) \\ Q(n) & Q(n-1) & Q(n-2) & ... & Q(n-m) \\ \sin(\theta(n)) & \cos(\theta(n)) & 0 & ... & 0 \\ |x(n)| & |x(n-1)| & |x(n-2)| & ... & |x(n-m)| \\ |x(n)|^2 & |x(n-1)|^2 & |x(n-2)|^2 & ... & |x(n-m)|^2 \\ & & \vdots & & \\ |x(n)|^q & |x(n-1)|^q & |x(n-2)|^q & ... & |x(n-m)|^q \end{bmatrix} \tag{4.5}$$
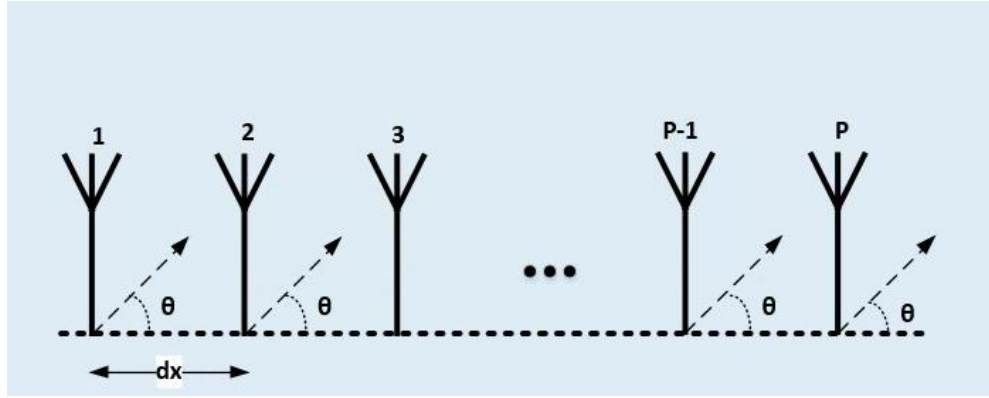
Figure 4.5 1D beamforming toward the direction θ.

To provide information about the steering angle, $\sin(\theta(n))$ and $\cos(\theta(n))$ of the beam steering angle θ are provided to the system. The angle information is The reason for using the pre-processed data in the model is to help reduce the model complexity while providing enough information to the CNN model. To understand the reason for using $\sin(\theta(n))$ and $\cos(\theta(n))$ in the input image matrix to the convolutional layer, the effect of beamforming weights on the transmitted signal needs to be studied. To simplify the mathematical analysis and without loss of generality, the following analysis considers 1D beamforming, to steer the beam toward the direction θ. Figure

4.5 shows the 1D beamforming toward the direction θ using P radiating elements. To steer the beam toward the direction θ, the pth radiating element is phase rotated using the following equation:

$$\alpha = (p - 1). 2\pi. \frac{dx}{\lambda}. \cos(\theta) \tag{4.6}$$

In a generic form the w_s is introduced to element p of the linear array antenna: [38]:

$$w_s = \{ 1, e^{j.2\pi.\frac{dx}{\lambda}.\cos(\theta)}, e^{2.j.2\pi.\frac{dx}{\lambda}.\cos(\theta)}, \dots, e^{(P-1).j.2\pi.\frac{dx}{\lambda}.\cos(\theta)} \} \tag{4.7}$$

Here θ is the steering angle, dx is the distance between two antenna elements and λ is the wavelength. Using equation (4.7), the pth antenna element, the signal driving the $(p + 1)^{th}$ PA is described as:

$$x_p(t) = x(t)\, e^{jp\alpha} = x(t)\, e^{-j2\pi.\frac{dx}{\lambda}.p.\cos(\theta)} \tag{4.8}$$

Equation (4.8) has the Jacobian form of $x(t)\, e^{jz.\cos(\theta)}$, with $z = -2\pi.\frac{dx}{\lambda}.p$. Using Jacobi-Anger expansion [46], $e^{jz.\cos(\theta)}$ can be described as follows:

$$e^{jz.\cos(\theta)} = \sum_{k=-\infty}^{\infty} j^k\, J_k(z)\, . e^{jk\theta} \tag{4.9}$$

where $J_k(z)$ is the kth Bessel function of the first kind. Now using equation (4.8) and (4.9), the input signal to the PA can be re-written as follows:

$$x_p(t) = x(t) . \sum_{k=-\infty}^{\infty} j^k\, J_k(z)\, . e^{jk\theta} \tag{4.10}$$

Using the Bessel function property, $J_{-k}(z) = (-1)^k. J_k(z)$, equation (4.10) can be re-written as follows:

$$x_p(t) = x(t) . (J_0(z) + 2. \sum_{k=1}^{\infty} j^k J_k(z) . cos(k\theta)) \tag{4.11}$$

Using the Binomial Expansion, $cos(k\theta)$ can be written as a linear combination of terms of the

form $\cos^{k-2r}(\theta) \times \sin^{2r}(\theta)$ as follows:

$$cos(k\theta) = \sum_{r=0}^{2r \leq k} (-1)^r \binom{k}{2r} \cos^{k-2r}(\theta) \times \sin^{2r}(\theta) \tag{4.12}$$

Using equations (4.11) and (4.12), the signal driving the pth PA can be written as follows:

$$x_p(t) = x(t) . (J_0(z) + 2. \sum_{k=1}^{\infty} j^k J_k(z) (\sum_{r=0}^{2r \leq k} (-1)^r \binom{k}{2r} \cos^{k-2r}(\theta) \times \sin^{2r}(\theta)) \tag{4.13}$$

To simplify the calculations and without loss of generality, we will consider in the following

analysis only the terms up to the third order Bessel function ($J_k(z)$, where $k < 4$. The signal

driving the pth PA can be re-written as shown in equation (4.14).

$$x_p(t) = x(t) . (J_0(z) - 2. J_1(z). \cos(\theta) + \cdots 2j. J_2(z). (\cos^2(\theta) - \sin^2(\theta)) -$$

$$2. J_3(z). (4. \cos^3 \theta - 3. \cos(\theta)) \tag{4.14}$$

Also for the sake of simplicity and without loss of generality, we consider a PA that can be modeled

using a 3rd order memoryless polynomial, where the PA output is given by:

$$y_p(t) = a_0 + a_1. x_p^1(t) + a_2. x_p^2(t) + a_3. x_p^3(t) \tag{4.15}$$

Using equations (4.14) and (4.15), the final PA output is given by:

$$y_p(t) = a_0 + a_1. (x(t) . (J_0(z) - 2. J_1(z). \cos(\theta) + 2j. J_2(z). (\cos^2(\theta) - \sin^2(\theta)) -$$

$$2. J_3(z). (4. \cos^3 \theta - 3. \cos(\theta))) + a_2. x^2(t)((J_0(z)^2 + 4. J_1(z)^2. \cos^2(\theta) -$$

$$4j. J_2(z)^2. (\cos^2(\theta) - \sin^2(\theta))^2 + 4J_3(z)^2 . (4. \cos^3 \theta - 3. \cos(\theta))^2) - 4J_0(z). J_1(z). \cos(\theta) +$$

$$4j.J_0(z).J_2(z).(\cos^2(\theta) - Sin^2(\theta)) - 4.J_0(z).J_3(z).\big((4.\cos^3\theta - 3.\cos(\theta)) -$$

$$8j.J_1(z).J_2(z).(\cos^3(\theta) - \sin^2(\theta).\cos(\theta)) - 8jJ_2(z).J_3(z).\big((4\cos^5(\theta) - 3\cos^3(\theta) -$$

$$4\sin^2(\theta).\cos^3(\theta) + 3\sin^2(\theta).\cos(\theta))\big)\big) + a_3.x^3(t) \dots \qquad (4.16\ )$$

The far-field received signal in the direction of θ, is the sum of all radiating element and the effect of channel and distance for each element. Equation (4.16) is the transmitted signal from each PA in a beamforming setup which is too complex to be used for modeling in practical applications as it contains different parameters. As discussed earlier, AI-based models are suitable for modelling complex mathematical representations. It is worth mentioning that, if it is not trained with the right information, AI model can not provide a competitive performance to other techniques. For AI model to offer good performance it should use the adequate information and a proper model with optimum number of neurons, filters and biases. As presented in equation (4.16), the PA model in beamforming is function of orders of $\sin(\theta(n))$ and $\cos(\theta(n))$ as well as the input signal. Thereby, to increase the modeling performance, $\sin(\theta(n))$ and $\cos(\theta(n))$ are added to the input image entries.

4.3.2 Convolutional Layer

The convolutional layer is composed of several digital filters. Figure 4.6 shows convolutional layer of the proposed AI-DPD architecture. The convolutional layer extracts the useful correlations between the input data. In the case of AI-DPD, it extracts the correlation between the beam direction and signal amplitude and phase. Equation (4.17) expresses the output of the dth filter CK_out_d as a function of f_d activation function, i^{th} input image $Image_i$, filter coefficients $Filter_d$ and filter biases $Bias\_filter_d$ of the dth filter. Here the CK_out_d is a matrix of size $W_{ck} \times L_{ck}$, $Image_i$ is a matrix of size $W_{in} \times L_{in}$, $Filter_d$ is a matrix of size $L_F \times W_F$ and

79
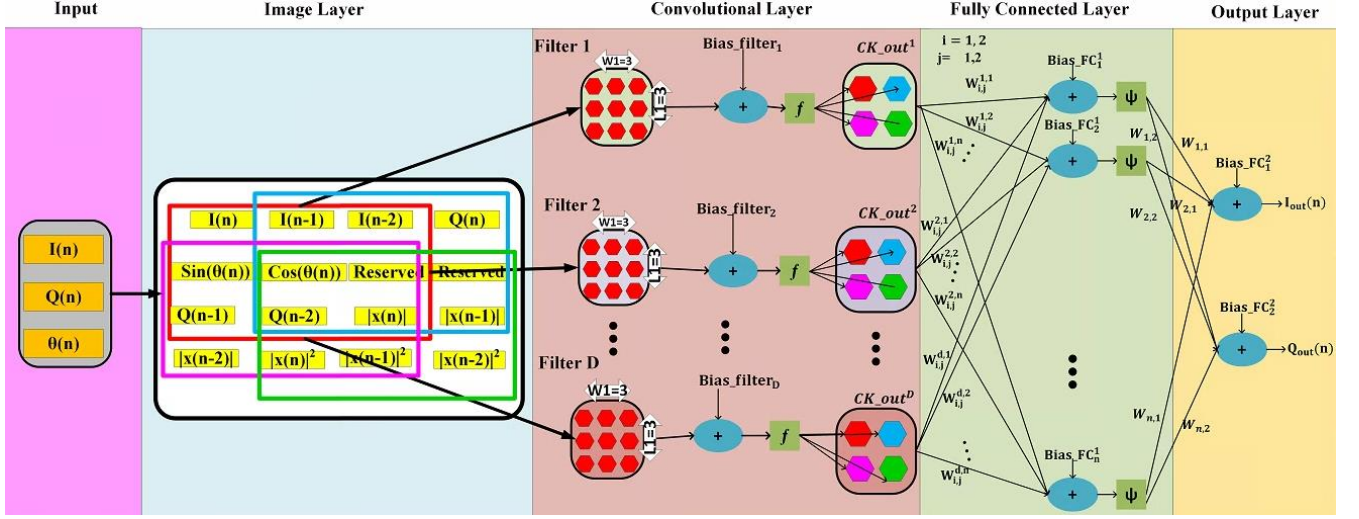
Figure 4.6 Architecture of the adopted CNN used in the developed AI-DPD.

$Bias\_filter_d$ is a scaler number. The operator $\otimes$ in equation (4.17) presents the convolution operation. In convolution operation, the filter slides over the input image and complete a dot operation and projects the result to the output pixel.

$$CK\_out_d = f_d \left(Image_i \otimes Filter_d + Bias_{filter_d} \right) \qquad (4.17)$$

The size and number of filters are system dependant, and they are determined using simulations. The designer should establish a trade-off between complexity and performance.

4.3.3   Fully connected layer

As illustrated in Figure 4.6, the output of the convolutional layer is connected to a fully connected neural network layer. The fully connected neural network is responsible for generating the pre-distorted signal. As expressed in equation (4.18), the feature maps (output of convolving filters) $CK\_out_d(i,j)$ are multiplied to connecting weights $W_{i,j}^{d,g}$ in the first layer. The important point here is that $W_{i,j}^{d,g}$ for $d = 1:D$ are equal, meaning the feature maps share the same weights

to connect to the fully connected layer [44]. This helps reducing the identification complexity.

Equation (4.18) expresses the output of the gth neuron in the first layer of the fully connected layer. The equations (4.19) and (4.20) express the pre-distorted signal $I_{out}(n)$ and $Q_{out}(n)$. Here, $Bias\_FC_g^1$ is the bias corresponding to the gth neuron in the first NN layer, $Bias\_FC_1^2$ and $Bias\_FC_2^2$ are the biases for the first and second neuron in the output layer of the NN. $W_{1,g}$ and $W_{2,g}$ are the connecting weights between the gth neuron in the first layer and two output neurons in the fully connect neural network layer.

$$FC\_out_g^1 = \psi \left(\sum_{d=1}^{D} \sum_{j=1}^{J} \sum_{i=1}^{I} CK\_out^d(i,j) \times W_{i,j}^{d,g} + Bias\_FC_g^1 \right) \qquad (4.18)$$

$$I_{out}(n) = \sum_{g=1}^{G} FC\_out_g^1 \times W_{1,g} + Bias\_FC_1^2 \qquad (4.19)$$

$$Q_{out}(n) = \sum_{g=1}^{G} FC\_out_g^1 \times W_{2,g} + Bias\_FC_2^2 \qquad (4.20)$$

4.4 Optimization

In order for the CNN model to offer the best modeling performance, it should be trained using adequate optimization settings. The used optimizer cost function is the Mean Square Error (MSE). The MSE at the iteration r is calculated as:

$$MSE(r) = \frac{1}{N} \sum_{n=1}^{N} \left[(I(n) - I'(n))^2 + (Q(n) - Q'(n))^2 \right] \qquad (4.21)$$

Here I', Q', I and Q are the modeled and measured I and Q data, respectively. N is the number of samples used to calculate the MSE. Gradient decay is an optimization algorithm that is widely used in artificial intelligence learning with different parameters. The modeling accuracy depends on how well a model is trained. There are different parameters affecting the identified model coefficients such as learning rate, number of Max Epochs, sample size, gradient decay factor and momentum. Learning rate and momentum determine how fast the algorithm converges toward the optimum parameter. MaxEpochs is the dataset size used in forward and backward training of the network. The dataset can't be used in training at once, it is divided to the number of batches with a defined sample size. To obtain the best modeling accuracy, the learning rate, number of MaxEpochs, sample size, gradient decay factor, and the momentum are swapped while modeling performance is recorded. The model is trained to minimize the MSE cost function. The MSE is calculated using equation (4.21). After the model is trained, the modeling performance is measured using the NMSE of the error between the desired value and model output. The NMSE is calculated using equation (2.2). The modeling performance for swapped parameters expressed in terms of NMSE are reported in Table 4.1. Table 4.1 reports the effect of the different training parameters on the modeling performance. Based on the results, the learning rate is selected to be 2.3e-3, MaxEpochs is 50, sample size is 550, gradient decay factor is 0.94 and momentum is 0.6. The other important parameter is the activation function. Table 4.2 reports the modeling accuracy of the CNN for different activation functions. According to the results, using Leakly ReLU activation function results in a 1.4 dB improvement in modeling performance compared to other activation functions.

Table 4-1 The modeling performance trained using different training parameters value.

Learning Rate	MAxEpochs	Sample Size	Gradient Decay factor	Momentum	Modeling Accuracy NMSE (dB)
2.3e-2	50	1000	0.91	0.50	-35.1
2.3e-2	40	550	0.94	0.60	-35.2
2e-2	35	400	0.95	0.65	-34.9
5e-3	50	1000	0.90	0.50	-35.0
5e-3	40	550	0.92	0.60	-34.9
5e-3	30	400	0.94	0.65	-34.1
1.8e-3	50	1000	0.90	0.50	-34.0
1.8e-3	40	550	0.93	0.60	-34.7
1.8e-3	30	400	0.95	0.65	-33.6
1.8e-3	25	250	0.93	0.67	-33.9
1e-3	50	1000	0.90	0.50	-33.7
1e-3	40	550	0.95	0.60	-34.1
5e-4	40	1000	0.91	0.55	-34.9
5e-4	30	550	0.93	0.60	-34.7
5e-4	25	250	0.95	0.68	-34.1

Table 4-2 The modeling performance of different activation functions.

Activation function	Modeling Accuracy NMSE (dB)
Relu	-34.8
elu	-34.2
Leakly ReLU	-35.2
Sigmoid	-33.8
Tanh	-34.4

4.5 Parameter selection

Here, the three important parameters are the number of convolution filters, the size of each filter, and the number of neurons in the fully connected layer. By increasing the number of neurons and filters, the modeling accuracy increases; however, the system is at the risk of over-
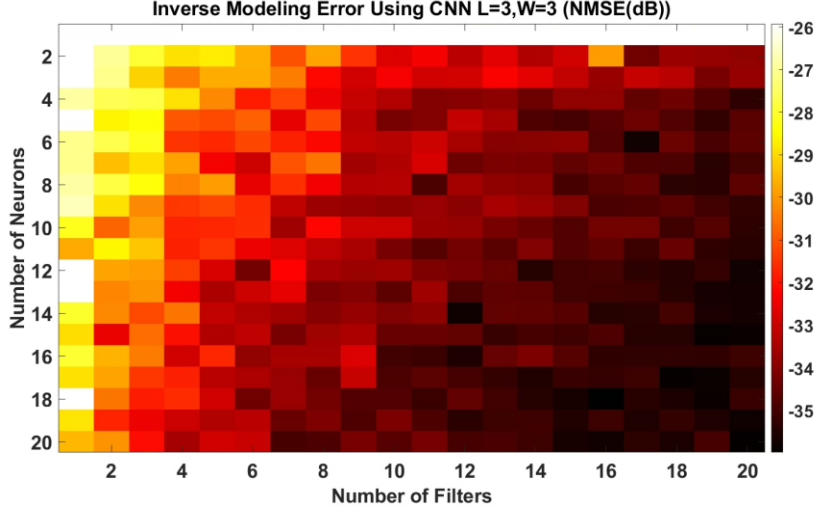
Figure 4.7 Inverse Modeling Accuracy of the proposed AI-DPD.

parameterization. Thereby it is important to find the optimum number of the CNN parameters. To do so, a series of MATLAB simulations have been carried out to determine the optimum value for the parameters identified earlier. Figure 4.7 shows the inverse modeling accuracy of the CNN versus the number of filters and neurons. Here, to establish a trade-off between complexity and performance, the number of filters and neurons are selected to be 17 and 6, respectively.

4.6   Complexity Analysis

The total number of learnable coefficients is an important factor to study any model's complexity. Suppose that the L_F is the length and W_F is the width of the convolution filters (CK), F is the number of filters in a single convolutional layer. The total number of filter coefficients, Q, is calculated as follows:

$$Q = (L_F . W_F) . F + F \tag{4.22}$$

Here, the second F is added to account for the number of bias weights at the output of each filter. The size of the output image of the convolution filter is a function of the padding size, input image size, and convolutional filter size. The $padding_{rows}$ and $padding_{rows}$ are extra columns and rows of zeros added around the data matrix. The size of the output image of the convolution layer is calculated as follows:

$[W, L] =$

$$[(W_{in} - W_F + 1 + 2.padding_{rows}), (L_{in} - L_F + 1 + 2.padding_{column})] \qquad (4.23)$$

Where the W_{in} and L_{in} are the input image size, width and length. The final outputs of CK layers are passed to a fully connected neural network to calculate the pre-distorted signal. The total number of learnable parameters in the fully connected neural network layer Z_{FC} is computed as follows:

$$Z_{FC} = (W \times L) \times N_1 + N_1 + N_1 \times N_2 + N_2 \qquad (4.24)$$

Here, N_1 and N_2 are the numbers of neurons in the first and second layer, respectively, of the fully connected neural network. In summary, the total number of learnable coefficients for the used CNN, including the parameters for the convolutional layer and fully connected neural network layer, are calculated as follows:

$$Learnables = Z_{FC} + Q \qquad (4.25)$$

The model complexity is studied from two perspectives: the number of coefficients and number of Floating-point Operations (FLOPs). Each FLOP is considered as one single floating-point operation like division, multiplication, subtraction, addition, exponentiation, etc. Although it takes the processor a different time to complete each task, they are considered as one FLOP.
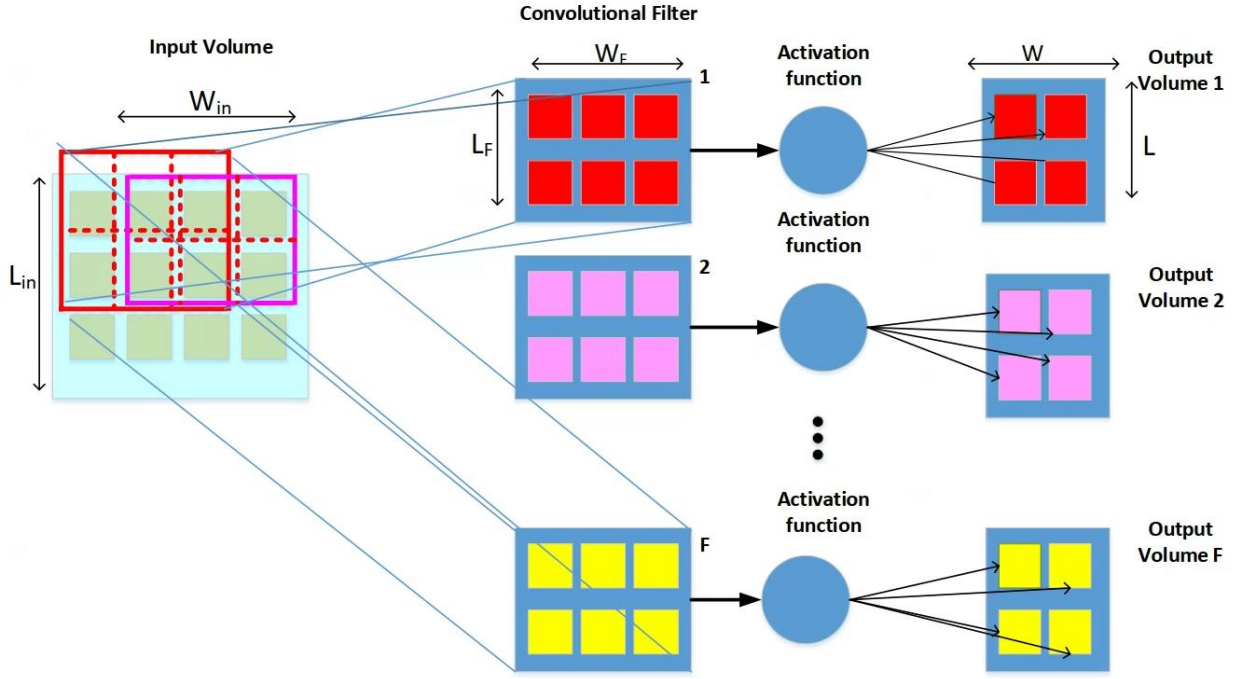
Figure 4.8 Convolutional filter operation process.

Thereby, to roughly estimate the FLOPs, the number of required mathematical operations per sample should be estimated. In CNN, the total number of FLOPS are the sum of FLOPs in the convolutional layer and the fully connected layer. Figure 4.8 presents the operation concept of the convolutional filters. The filter slides over the input pixels and performs L_F . W_F dot operation, and projects the results to the output pixels. Same operation is repeated to generate all the output pixels. Therefore, the number of multiplications in convolutional layer $multiplications_{CK}$ is calculated as follows:

$$multiplications_{CK} = L_F . W_F . (W . L) . F \qquad (4.26)$$

Every time the filter slides over the input L_F . $W_F - 1$ addition and a bias added to this value is performed in the activation layer. The same operation is repeated to generate the output

volume. Therefore, the number of addition operation in the convolutional layer is calculated as follows:

$$addition_{CK} = (L_F . W_F) . (W . L) . \text{F} \tag{4.27}$$

The output of each filtering iteration passes through an activation function, as a result, the number of FLOP required for activation functions are estimated as:

$$AF_{ck} = F . (W . L) \tag{4.28}$$

Here, F is the number of convolution filters. Similarly, the number of multiplications and additions in the fully connected neural network layer is estimated using the following equations:

$$multiplications_{FC} = N_1 . (W . L) + N_1 . N_2 \tag{4.29}$$

$$addition_{FC} = (W . L + 1) . N_1 + (N_1 + 1) . N_2 \tag{4.30}$$

The activation functions also take up processor time and they need to be accounted for the in FLOPS calculation. The number of activation functions in the fully connected neural network layer are calculated as follows:

$$AF_{FC} = N_1 + N_2 \tag{4.31}$$

Using equations (26) to (31), one is able to estimate the total number of FLOPs per sample in the CNN including the convolutional layer, the FC layer and the output layer.

$$FLOPs = multiplications_{CK} + multiplications_{FC} + addition_{CK} + addition_{FC} + (AF_{CK} +$$

$$AF_{FC}) \times H \tag{4.28}$$

Here H is the number of required FLOPs per activation function.

4.7   Measurement Results

Figure 4.9 shows the block diagram of the developed AI-DPD verification setup. Different components of the setup are thoroughly described in chapter 3. Figure 4.10 shows a photograph of the setup used for AI-DPD algorithm verification. The input data-steering angle and modulated data- are processed and organized in a 2D format and passed to the AI-DPD actuator. The DPD coefficients are calculated offline using MATLAB. Then the data passes through the transmission path and is transmitted using the 16-element phased array antenna at the 3.5GHz center frequency. The far-field signal is received 3 meters away and then it is phase and time-aligned for further analysis. As discussed earlier, the transmitter is a MEGAbeecube composed of 4 AD9361 SDR transceivers and Zynq 7000 series, frequency and time-synchronized using the techniques described in Chapter 3.

To gain a better understanding of the transmitter behaviour, the beam is steered across a steering range from -50° to 50° with 5° steps while the PAs are transmitting at average power of 21 dBm, operating in room temperature and at the center frequency of 3.5 GHz. The far-field received signal is sent to PC for further analysis. For AI-DPD training 5 angles at {-50°, -24°, 0°, 29°, 58°} are selected to train the model. At each angle 10k data samples are used in the training matrix. Therefore, 50k data samples are used in one epoch to train the network. A total of 40 epochs are used to train the model which achieves to -36.3 dB NMSE. To validate the model 5 different angles at {-35°, -14°, 7°, 37°, 47°} are used. The validation accuracy described in the term of NMSE is -35 dB. To compare the complexity and modeling accuracy of different DPD models described in Chapter 2, different models are used to linearize the developed beamforming transmitter. Table 4.3 describes each DPD model architecture. The Bo-DPD [24] is a directional
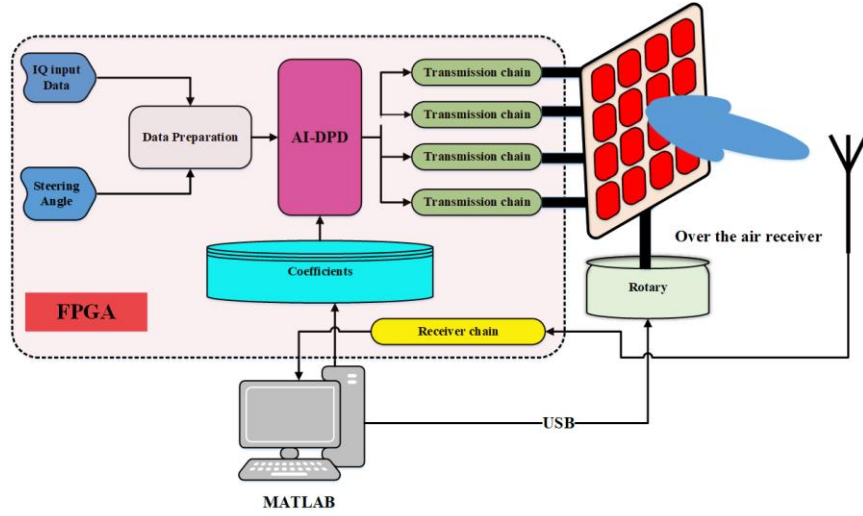
Figure 4.9 Proposed CNN-DPD architecture using 4x4 phased array transmitter antenna and far-field over the air signal acquisition antenna.
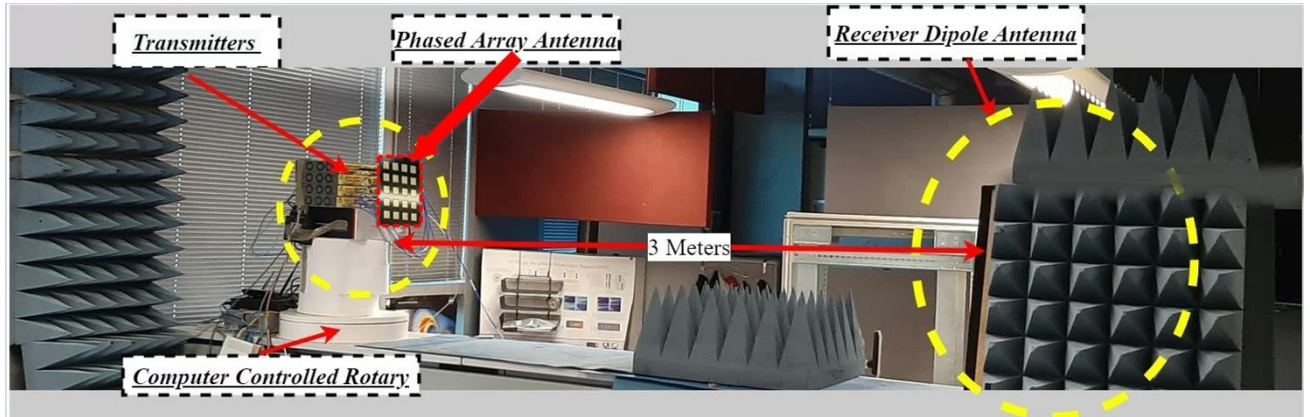
Figure 4.10 Experimental setup developed for MIMO digital beamforming and algorithms verification.

DPD requiring the DPD coefficients to be updated at each steering angle. Bo-DPD coefficients are extracted at a specific direction so they can linearize the transmitter within a limited angle width around the extracted point; this range is called a cluster. This technique is used in [45] to reduce computational complexity.

In [45] the steering range is divided into a number of clusters and a memory polynomial is extracted to linearize the beam in that direction. The number of clusters depends on various

89

Table 4-3 Different DPD architecture models used for comparison.

DPD	Model	Settings	
CNN- With sinθ and cosθ	CNN	Model	CNN
		Image size	4x4
		Convolving Filter	3x3
		Kernel size	17
		FC neurons	6-2
		Activation Function	LeakyReLu
		Memory depth	2
CNN- With θ		Model	CNN
		Image size	4x4
		Convolving Filter	3x3
		Kernel size	22
		FC neurons	9-2
		Activation Function	LeakyReLu
		Memory depth	2
CNN-no angle	CNN	Model	CNN
		Image size	3x4
		Convolving Filter	2x3
		Kernel size	20
		FC neurons	8-2
		Activation Function	LeakyReLu
		Memory depth	2
RVTDNN With sinθ and cosθ	DNN	Neurons in each layer	(14-37-12-2)
		Activation function	Tanh
		Memory	2 previous data
RVTDNN with θ	DNN	Neurons in each layer	(13-49-18-2)
		Activation function	Tanh
		Memory	2 previous data
DNN [28] No angle	DNN	Neurons in each layer	(12-30-10-2)
		Activation function	Tanh
		Memory	2 previous data
BO-DPD [24]	Memory polynomial	Order=14, memory tap=2, Least square to estimate the coefficients using 15000 samples	
[45]	Memory polynomial	Order=15, memory tap=2 Least square to estimate the coefficients using 15000 samples	

parameters such as beam width and frequency. In our case, the steering range is divided into 9

clusters and the DPD coefficients are calculated within each cluster. For a fair comparison between the different DPD techniques, NN and CNN are used to model the transmitter using steering angle θ, sin(θ) and cos(θ), and without any angle information. Table 4.4 describes each model's complexity analysis with respect to model parameters. These equations are used to estimate the DPD models complexity reported in Table 4.5. Table 4.5 compares each DPD architecture

Table 4-4 complexity Analysis of different techniques

Characteristic \ Model	Number of coefficients		Flops/sample
BO-DPD [24]	$(M+1).K$		$8.((M+1).K)-2$
	M: memory depth, K: nonlinearity order		
Full Angle DPD [27]	$(S.(M+1).K) + n_f.((S.(M_f+1).K_f))$		$8.((M+1).K)-2 + n_f(8.((M_f+1).K_F)-2)$
	M: memory depth, K: nonlinearity order, n_f number of fine-tuning boxes, M_f:memory depth and K_f: nonlinearity order of fth fine tuning box		
Shallow Neural Network [8]	$(N_i+1).N_1 + (N_1+1).N_o$		$2.N_i.N_1 + 2.N_1.N_o + 13N_1$
	N_i, N_1, N_o: number of neurons in input, middle and output layer		
Deep Neural Network [28]	$(N_i+1).N_1 + \sum_{f=2}^{F}(N_f+1).N_f + (N_F+1).N_0$		$2.N_i.N_1 + 2.N_F N_0 + 2.\sum_{f=2}^{F} N_{F-1}.N_f + 15.\sum_{f=1}^{F} N_f$
	N_f: number of neurons in f^{th} layer		
CNN	Without transfer technique	$L_F.W_F.F + W.L.N_1 + N1 + N1.N2 + N2$	$F.W.L.N1 + N1.N2.N1 + \cdots + (W.L).N1 + (N1+1).N2 + \cdots + (F.W.L + N1 + N2).H$
	Using Transfer technique	$(N_1+1).N_1 + (N_1+1).N_2$	
	W,L: width and length of the output image of the convolving layer, W_F, L_F: Width and Length of the convolutional filter, F: number of convolving filter, N1 and N2: neurons in the first and second layer of NN, N_i and N_o: neurons in the input and output layer of NN H: flops for activation function.		

modeling accuracy and their features. According to the results reported in Table 4.5, the inclusion of sin(θ) and cos(θ) reduces the system complexity while increasing the modeling performance. Moreover, the CNN-based DPD offers 1.2 dB improvement in modeling performance in comparison to DNN.

Table 4-5 The architecture, performance, and complexity analysis of different techniques.

	AI-DPD Using sin(θ) and cos(θ)	AI-DPD Using θ	CNN without steering angle information	RVTDNN Using sin(θ) and cos(θ)	RVTDNN Using θ	DNN [28]	BO-DPD [24]	DPD [45]
Direction dependency compensation	✔	✔	✔	✔	✔	✔	✔	✔
I/Q imbalance compensation	✔	✔	✔	✔	✔	✔	×	×
Number of training angles per subarray	5	5	17	5	5	17	13**	9 clusters
Average Modeling Accuracy (NMSE-dB)	-35. 3	-34.2	-32.9	-34.3	-33.5	-32.7*	-35.8	-32.9
Complexity (learnable parameters)	598	1041	806	726	1066	522*	39×13=507**	45×9=405***
FLOPs/sample	1268	1940	1442	2107	2999	1560*	334**	358***

*Implemented using the model presented in [28]

** Implemented using approach presented in [24]

*** Implemented using approach presented in [45]

The model complexity is studied in terms of FLOPS and number coefficients. Depending on the implementation platform, the complexity can be translated into processing unit. In the next step, the performance of each DPD model is verified using the designed MIMO beamforming platforms. Figure 4.11 compares the AM/AM and Power Spectral Density (PSD) of the transmitted signal linearized using different techniques. The transmitted signal with no-DPD is compressed by 4 dB with an average NMSE of -21 dB, average ACPR of -31 dBc across the steering range. As shown in Figure 4.11, the Bo-DPD and the AI-DPD are both able to linearize and compensate for
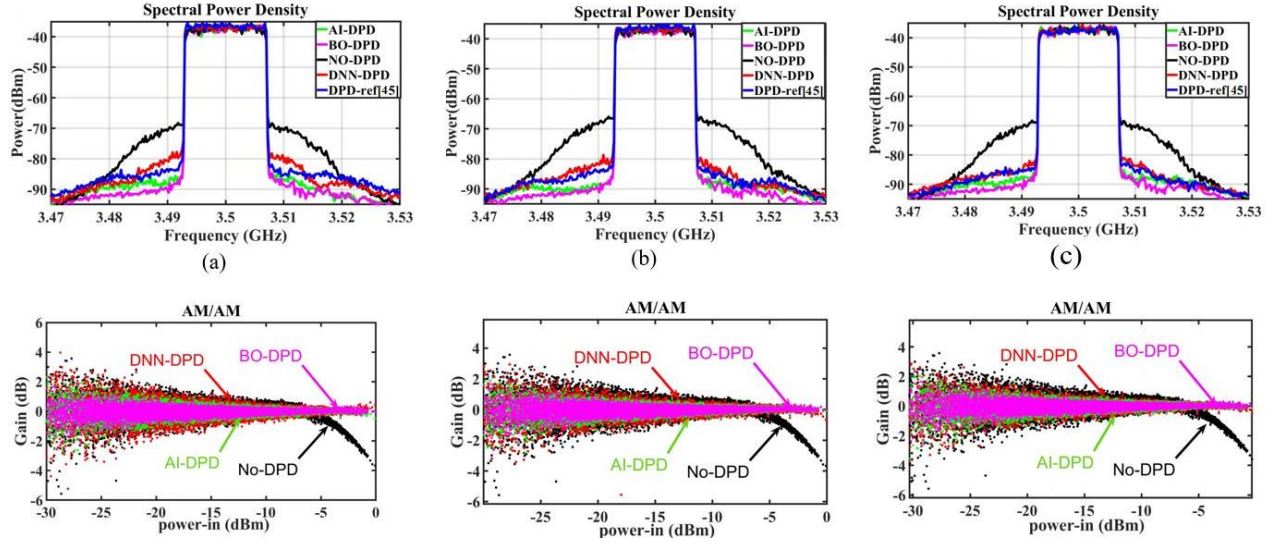
Figure 4.11 PSD of the received signal at: (a). 30°, (b). 10°, (c). -30°, AM/AM at: (d). 30°, (e). 10°, (f). -30°.

distortions in a similar manner. Figure 4.11 (d) (e) and (f) shows the AM /AM of the transmitted signal. The results show that AI-DPD and Bo-DPD can both substantially reduce the dispersion while there is more dispersion left when using other techniques. Figure 4.12 shows the ACPR and NMSE of the transmitted signal using different techniques across the steering range.

Again, the Bo-DPD is extracted at each steering direction. As the direction changes, the Bo-DPD coefficients should be updated. This approach provides the best-case scenario in terms of linearization performance since a dedicated DPD is used for each direction. The proposed AI-DPD offers a close linearization performance in terms of ACPR and NMSE when compared to the Bo-DPD. Indeed, its average of 48.6 dBc ACPR across the steering range is close to the -49.1 dBc ACPR for the Bo-DPD. Similarly, the AI-DPD achieved an average NMSE equal to -34.3 dB which is close to the Bo-DPD NMSE of -35.5 dB. It is worth noting that the DPD in [45] offers a close linearization performance to the DNN DPD. The reason is that both DPD architectures use the averaging technique, which sacrifices the performance to reduce the computational complexity.

4.8   Conclusion

It was shown that the AI-DPD linearizes the beamforming transmitter based on the steering angle using a limited number of training angles. The AI-DPD provides a good linearization performance across the steering range and eliminates the need for a dedicated observation path. Removing the observation pass substantially reduces the hardware complexity and manufacturing cost in the massive MIMO base stations. The measurement results show that the proposed AI-DPD
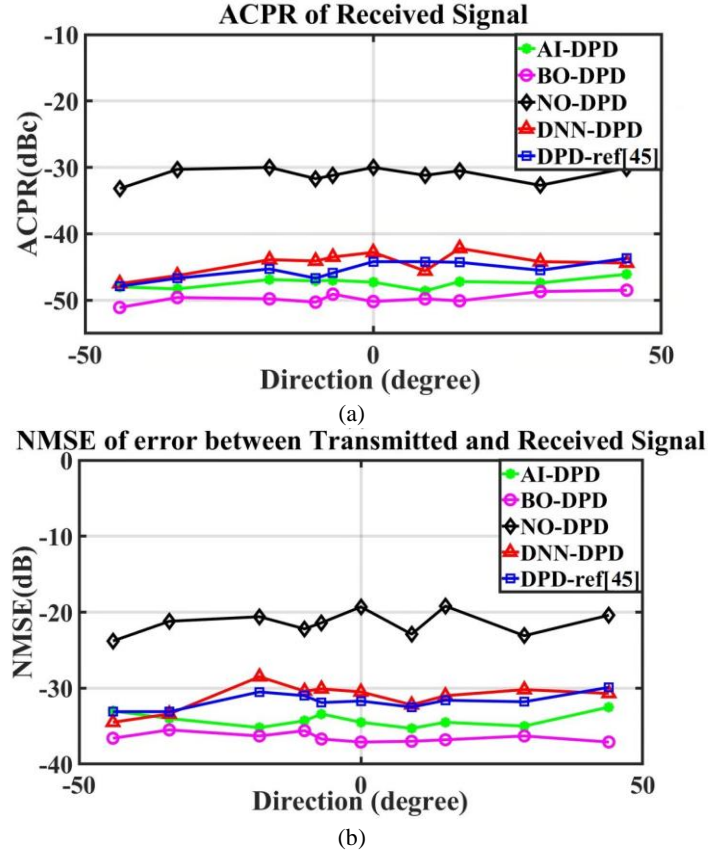
Figure 4.12 ACPR of the received signal linearized using different methods, (b) NMSE of the received signal.

provides similar linearization performance in comparison to the Bo-DPD while not requiring the

DPD adaption at each steering angle.

5 Chapter 5: Power and Temperature Inclusive Digital Pre-Distortion

5.1   Introduction

The power amplifiers in the base stations are often pushed to operate near saturation regions to improve the base station's power efficiency. However, operating near saturation region introduces nonlinearities and distortions to the original signal that needs to be compensated using DPD.

Normally, the base stations are expected to operate at different signal's average power for different purposes like beamforming or when the user's distance from the base station is changing. Different operating power can significantly affect the PA nonlinear behaviour. The PA behaviour can be described using static and dynamic distortions. The static distortion is mainly influenced by a signal's power level and the base plate temperature. The dynamic behaviour is affected by the signal bandwidth and instantaneous temperature variation of the transistor's junction [48].

There are three main sources of the thermal variation: base plate temperature [49], signal's average power [49] and self-heating phenomena [50]. Self-heating is problematic in narrow band signals in the range of Kilohertz. The rate of signal amplitude change in a narrow bandwidth signal is slower than in a wider bandwidth signal. Therefore, the transistor operating point varies slower when compared to a wide bandwidth signal. This causes the transistor operating point as well as the transistor junction temperature to constantly fluctuate. This is a serious problem in narrow band signal used in 2G, and an in-depth study of this issue is out of the scope of this thesis. On the other hand, the rate of amplitude change in wide bandwidth signals (>1 MHz) are fast enough that transistor junction temperature is a function of the signal's average power which is the main focus of this chapter.

Radio base stations are installed in different geolocations where the ambient temperature may vary from -40° to +50° Celsius. Such a wide range of the ambient temperature variation negatively affects the PA behaviour and the DPD performance used to mitigate these distortions.

Normally, high power PAs are connected to a heat sink to increase the heat dissipation rate. The heat dissipation itself depends on the base plate temperature. Therefore, base plate temperature is another important factor that needs to be considered in DPD design.

In summary, the PA behaviour is a function of transistor junction temperature. However, the junction temperature is not accessible. For wideband signals, the junction temperature is a function of signal's average power and base plate temperature. Therefore, these two parameters must be taken into account when designing a robust DPD architecture for practical applications. To maintain the signal quality using conventional DPD techniques, the DPD coefficients need to be re-identified with any variation in the signal's average power or base plate temperature. In this case, a dedicated observation path is required to monitor the signal quality during the runtime, which increases the hardware complexity. In addition, the DPD coefficients re-identification needs to be fast enough to meet the timing requirements. Since the maximum allowable latency is 1 ms, the DPD should be adapt itself with new condition in less than 1 ms [23]. To address these issues, this chapter introduces a novel Power Temperature Inclusive DPD (PTI-DPD) to provide continuous linearization under the signal's average power and base plate temperature's varying conditions. The PTI-DPD is implemented using CNN to provide the best modeling performance. The proposed technique is extendible to cover more parameters like reflection coefficients, operating frequency, bandwidth, etc. However, this chapter only focuses on the signal's average power and base plate temperature. The rest of the chapter is organized into 3 sections. Section 2 introduces the PTI-DPD architecture. Section 3 studies the architecture modeling performance.

Section 4 outlines the effect that signal's average power and base plate temperature variation have on the transmitter's behaviour.

In summary, the main contributions of this chapter are:

1. Studying the effect of base plate temperature on the PA's behavior

2. Studying the effect of signal's average power on the PA's behavior

3. Introducing a novel DPD behavioral model for compensation of the PA nonlinearities in the presence of varying signal's average power and base plate temperature

5.2   Power and Temperature Inclusive Digital Pre-distortion

Traditionally, the PA behaviour is described using an independent memory polynomial at each PA's operating conditions. A general representation of the PA output $y(n)$ under a given base plate temperature, T, and a signal's average power, P, is as follows:

$$y(n) = \sum_{m=1}^{M} \sum_{k=1}^{K} a_{m,k}^{T,P} x(n-m).|x(n-m)|^{k-1} \tag{5.1}$$

Here $x(n-m)$ is the mth delay version of the input signal to the PA $x(n)$, the $a_{m,k}^{T,P}$ are the DPD coefficients for the mth memory order and k^{th} nonlinearity at the base plate temperature T and the signal's average power P. To model the transmitter using equation (5.1), the $a_{m,k}^{T,P}$ coefficients need to be identified for each power level and base plate temperature. DPD coefficients' identification should be repeat at different base plate temperature and power levels. Next section

lists the problems and possible solutions for wireless transmitters operating at different signal's average power and base plate temperature.

5.2.1   Problem Definition

1. The IQ imbalance are an inevitable imperfection in the commercial RF transceivers and have been discussed in chapter 3. IQ imbalance negatively impacts the signal quality. To compensate for the IQ imbalances, any model must be extracted using the input signal, complex conjugate of the input signal, and the output signal.

2. The PA behavior varies at different signal's average power and base plate temperature. Here, we assume that if a signal is wide enough, then the self-heating phenomena is negligible and only signal's average power and base plate temperature are the important factors here.

3. By increasing the number of transmitter settings and environmental parameters affecting the PA behavior, it is difficult to model the system for all possible scenarios using equation-based behavioral models.

Considering the issues detailed above, this chapter introduces a PTI-DPD implemented using CNN. The proposed PTI-DPD considers the base plate temperature and signal's average power and compensate for distortions added due to any variation of the parameters outlined previously. In addition, the PTI-DPD is modeled using a limited number of training conditions to reduce the training complexity. Since the input image contains in-phase and quadrature parts of the input signal, it can model the complex conjugate and compensate for the IQ imbalance. As it is shown
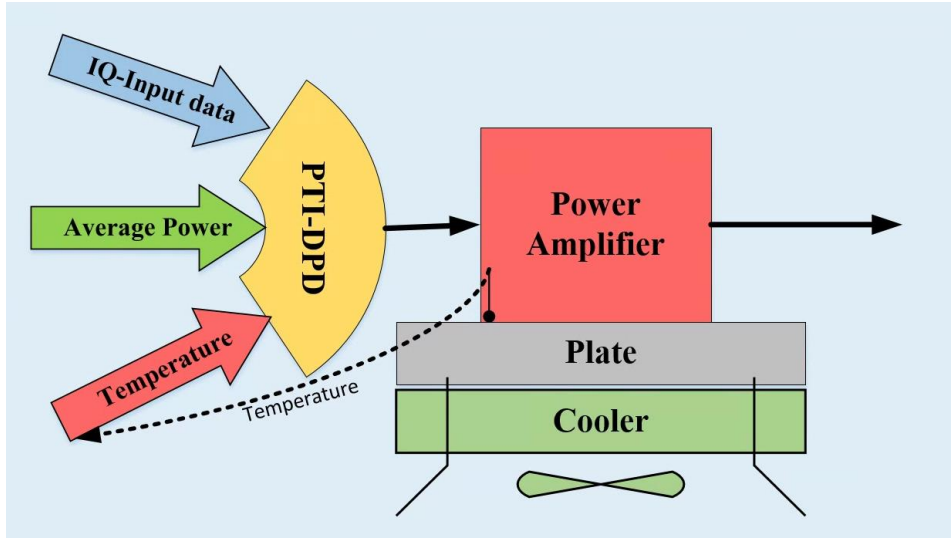
Figure 5.1 Top view of the proposed PTI-DPD [51].

in [49] the CNN can also compensate for IQ imbalance in the transmitter. Figure 5.1 depicts a top view of the proposed DPD architecture. The inputs to the system are the modulated signal, base plate temperature captured as close as possible to the PA substrate, and the signal's average power.

5.3   PTI-DPD architecture

Figure 5.2 illustrates the architecture of the CNN model used in PTI-DPD. The PTI-DPD model is composed of the input image layer, convolutional layer and a fully connected NN layer. The following paragraphs introduces CNN layers used in PTI-DPD.

5.3.1   Input image layer

The inputs are the signal's average power $\overline{P(n)}$ and base plate temperature $T(n)$ to cover one transmitter setting and one environmental factor in the input image. The input image also contains in-phase I(n), quadrature Q(n) of the input modulated signal and as well as the m delayed
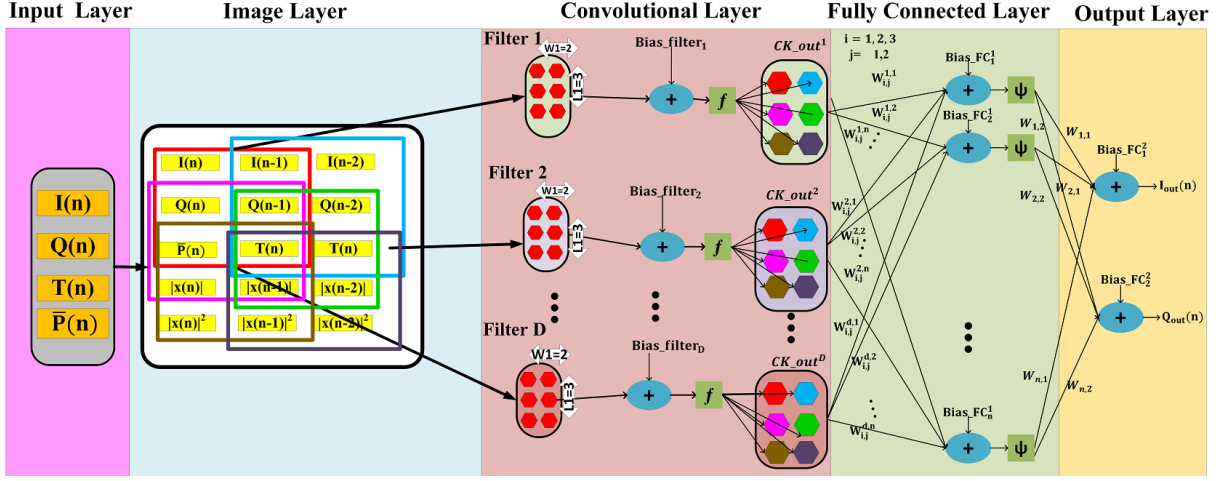
Figure 5.2 The architecture of the proposed RVTDCNN to implement PTI-DPD [51].

version and k order of the input signal envelope $x(n)$. The $\overline{P(n)}$ and $T(n)$ are located in the middle of the input image so that their effect is captured in more feature maps.

$$Image_i = \begin{bmatrix} I(n) & I(n-1) & I(n-2) & ... & I(n-m) \\ Q(n) & Q(n-1) & Q(n-2) & ... & Q(n-m) \\ \overline{P(n)} & T(n) & 0 & ... & 0 \\ |x(n)| & |x(n-1)| & |x(n-2)| & ... & |x(n-m)| \\ |x(n)|^2 & |x(n-1)|^2 & |x(n-2)|^2 & ... & |x(n-m)|^2 \\ & & \vdots & & \\ |x(n)|^k & |x(n-1)|^k & |x(n-2)|^k & ... & |x(n-m)|^k \end{bmatrix} \tag{5.2}$$

5.3.2 Convolutional layer

This layer is responsible for extracting the input image features using a set of convolution filters applied to the input image. The output of the filters is passed to the fully connected layer. The size and number of the convolution filters are determined using MATLAB simulations to achieve a trade-off between complexity and model performance.

5.3.3 Fully connected layer

The data at the output of convolution filters are converted into a 1-D array and passed to the fully connected layer. This layer is responsible for generating the pre-distorted signal.

5.3.4 Model extraction and training

The model parameters including the filter coefficients, and the neurons' biases and weights of the fully connected layer are trained to minimize the Mean Square Error (MSE) as expressed below:

$$MSE = \frac{1}{N} \sum_{n=1}^{N} \left[\left(I(n) - I'(n)\right)^2 + \left(Q(n) - Q'(n)\right)^2 \right] \tag{5.3}$$

And the modeling accuracy is analyzed in terms of NMSE as described in equation (2.2). Here, $I(n), I'(n)$ are the in-phase components of the desired signal and the modeled one, respectively. Similarly, $Q(n)$ and $Q'(s)$ are the quadrature components of the desired signal and the modeled one, respectively. To select the best optimizer parameters, learning rate, Maxepochs, sample size, gradient decay factor, and momentum have been given different numbers while monitoring the modeling performance. The mentioned optimizer parameters are defined in chapter 4.4. Table 5.1 reports the DPD modeling accuracy in term of NMSE for different model parameter settings. Based on the results, the DPD model with a gradient decay factor of 0.93, an initial learning rate of 1e-3, momentum rate of 0.6, and Maxepochs size of 25 provides the best modeling performance.

Table 5-1 Effect of optimizer parameters on the modeling performance [51]

Learning Rate	MaxEpochs	Sample size	Gradient decay factor	Momentum	Modeling Accuracy (NMSE-dB)
1e-2	40	1000	0.93	0.6	-37.8
1e-2	20	500	0.91	0.5	-36.8
1e-3	25	500	0.93	0.6	-37.7
1e-3	25	1000	0.93	0.6	-38.5
1e-3	25	2000	0.93	0.6	-38.2
1e-3	20	1000	0.93	0.6	-38.1
1e-3	25	1000	0.89	0.6	-38.2
5e-2	20	500	0.93	0.5	-37.2
5e-3	25	500	0.93	0.6	-36.7
5e-3	30	1000	0.93	0.5	-38.1
5e-3	20	2000	0.93	0.4	-37.5
5e-3	25	1000	0.93	0.6	-36.7
1e-4	25	1000	0.93	0.63	-38.1

@2021 IEEE

5.3.5 Data preparation

For the CNN to be able to extract the transmitter behaviour, it should be trained using the data that contains enough information describing the system behaviour. To do so, both signal's average power and base plate temperature have been swept and the system's behaviour is recorded. The input power has been backed-off from 0 to 8 dB to change the signal's average power. The Input Back-Off Power (IPBO) is calculated as follows:

$$IPBO = P_{in,Sat} - P_{in,peak} \qquad (5.4)$$

Here the $P_{in,Sat}$ is the input saturation power of the power amplifier and the $P_{in,peak}$ is the input signal peak power. Similarly, the temperature has been swept from 26° to 70° Celsius at different power levels and the transmitter behaviour is recorded. As previously mentioned, one of the main

advantages of the PTI-DPD over other techniques is that it does not need to be trained using all the possible combinations of the model inputs, which are the signal's average power, base plate temperature and input modulated signal. Here, a 3×3 matrix of data representing 3 base plate temperature levels at {24° ,48° ,70°} Celsius each captured at 3 input back-off levels: {0, 3.5, 7.5} dB back-off is used for PTI-DPD training. A total of 90 K samples are used to train the model in 50 epochs. The model achieves the modeling accuracy of -39.5 dB. To validate the model a total of 90 K samples captured at {0.5 2.5 6} back of level and at {27° ,38° ,69°} are used. The model achieves a training accuracy of -38.7 dB. It is worth mentioning that, to provide the model with adequate information, the input data used for training should cover the full range of the input parameters. In other words, it should include the maximum and minimum values of the input parameters.

5.4 Measurement setup

Figure 5.3 shows the experimental setup used for DPD model verification. The RF data is generated using AD9361 SDR platform. The SDR operates as 61.44 MHz sampling frequency and operating frequency of 2.14 GHz. The RF input signal is amplified using ZHL-42X driver to increase the power of the signal to the desired value to drive the main Doherty PA. The ZHL-42X has a gain of 38 dB and the Doherty PA has a 54.7 dBm saturation power at the operating frequency of 2.14 GHz. The 15 MHz signal is wide enough to trigger both dynamic and static behaviour. It is worth mentioning that, the power dependant parameters of the PA are considered as static characteristics, which can be extracted using narrow band signals, and further increase of the signal bandwidth does not affect the static behaviour.

The proposed PTI-DPD is compared with three other DPD configurations. The first configuration is called in this thesis AP-DPD. In this technique, the CNN model coefficients are calculated over the data captured at all average power and ambient temperature conditions. The only input to the AP-DPD model are the IQ signals with the same memory depth and nonlinearity
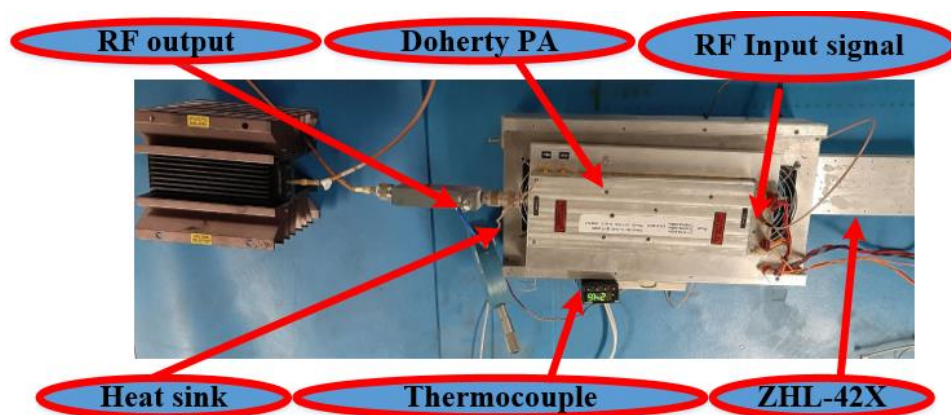
@2021 IEEE

Figure 5.3 Experimental setup used to validate the DPD techniques [51]

order as PTI-DPD. There is no information in regard of the average power and ambient temperature in AP-DPD. The other configuration is called in this thesis Max-P DPD. In this configuration, the CNN model coefficients are extracted using the data captured at the maximum output power corresponding to IPBO=0 and base plate temperature of 27° Celsius. The third configuration is using the reference DPD in [15]. The reference DPD [15] is a power adaptive DPD implemented using power adaptive memory polynomial. In this technique, a memory polynomial is extracted in parallel with linear and nonlinear correction based on the signal power as described in chapter 2.1.

5.5 Modeling performance

The CNN model is trained using the 3×3 data matrix. The input image to the system is a 4×3 pixel data that includes signal's average power, base plate temperature, and input IQ signal with 2 memory depth. To select the optimum number of neurons and filters, the number of neurons and filters are swept from 1 to 40 while monitoring the modeling performance. The optimum performance was obtained for 7 neurons is 7 and 12 filters with size of 3×2. To choose the best activation function, the best achievable modeling accuracy of PTI-DPD using different activation functions is calculated and shown in Table 5.2. The results show that Tanh, LeakyReLu and ReLu offer relatively similar performance. Since ReLu has the lowest implementation complexity compared to others, it is used as the activation function in the PTI-DPD.

Figure 5.4 shows the inverse modeling performance of the PTI-DPD versus IPBO and base plate temperature. The IPBO ranges from 0 to 10 dB and base plate temperature is swept from 27° to 69° Celsius. The results show that, on average, the PTI-DPD offers -37.2 dB modeling accuracy across the swept parameters.
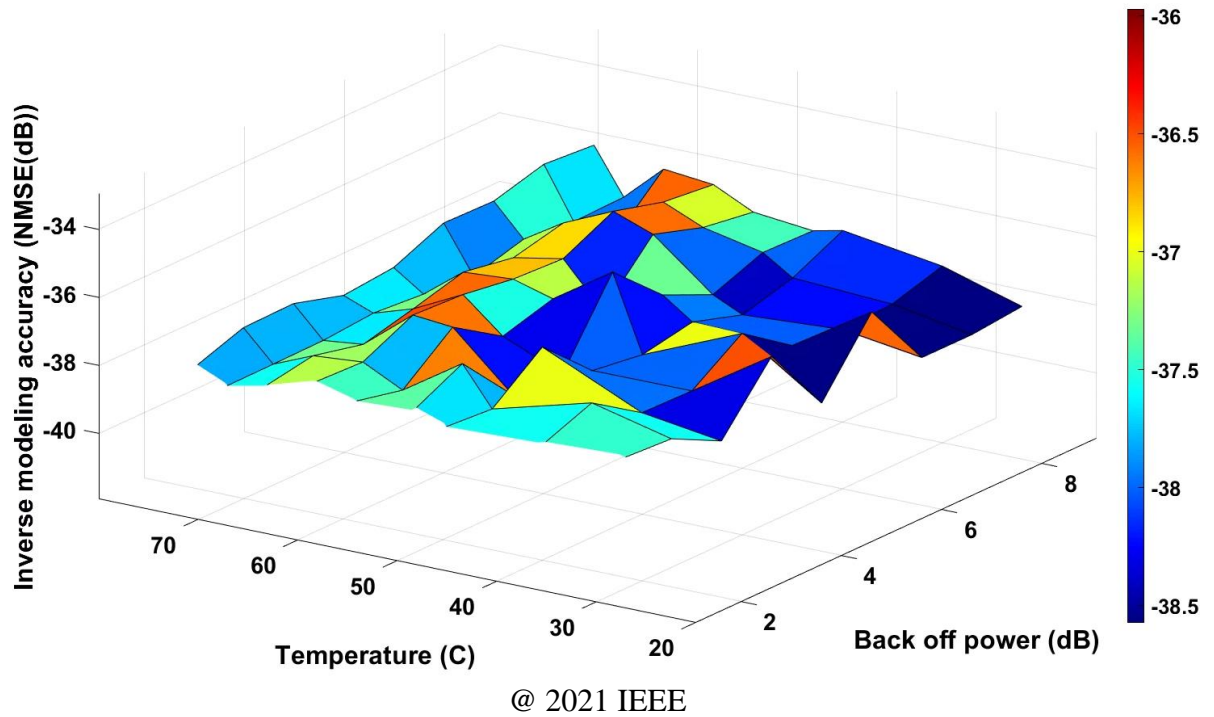
@ 2021 IEEE

Figure 5.4 Inverse modeling performance of the developed PTI-DPD [51]

Table 5-2. Modeling performance of the activation functions [51].

Activation function	Tanh	ReLu	Sigmoid	Leaky Relu
Modeling Accuracy (NMSE-dB)	-38.8	-38.6	-37.3	-38.2

@ 2021 IEEE

Table 5-3.Performance and complexity analysis for different DPD architecture [51].

Model property		PTI-DPD	Maxp-DPD	AP-DPD	Reference [15]
Power adaptation		✓	×	×	✓
Temperature adaptation		✓	×	×	×
I/Q imbalance compensation		✓	✓	✓	×
Best Modeling Accuracy (NMSE-dB)	IQI= 0°, 0 dB	-38.5	-38.6	-33.5	-36.1
	IQI= 3°, 1 dB	-38.3	38.1	-33.8	-33.7
Worst Modeling Accuracy (NMSE-dB)		-36.3	-	-28.4	-
Average Modeling Accuracy (NMSE-dB)		-37.2	-	-30.6	-
Learnable Coefficients		443	233	689	180
FLOPs/sample		840	500	1238	602
Image size		5×3	4×3	4×3	-
Convolution Filter		12	8	17	-
Kernel size		3×2	3×2	3×2	-
FC neurons		7	5	8	-
Activation Function		ReLu	ReLu	ReLu	-

@ 2021 IEEE

Different DPD models are analyzed in term of complexity, modeling performance and their capabilities in Table 5.3. The complexity is studied in terms of FLOPs and learnable coefficients using the relations described in chapter 4. The PTI-DPD and reference [15] are valid for different average power level, while AP-DPD and Max-P DPD are not. The PTI-DPD is valid for different base plate temperature without the need to re-identify the DPD when the base plate temperature

changes. However, the other DPD techniques do not possess such capability. In terms of modeling accuracy, the PTI-DPD provides a modeling accuracy equal to -38.5 dB while the AP-DPD provides only -33.8 dB. The reference DPD from [15] has a modeling accuracy of -36.1 dB.

In the following sections, the PTI-DPD behaviour is studied in three steps. The first step is to vary the average power only without changing the ambient temperature and measure the base plate temperature. In this case the temperature information is not feed to the PTI-DPD. The second step is to emulate the ambient temperature change by changing the fan speed. In the third step, both the average power and base plate temperature are varied and the performance of different DPD techniques are studied.

5.5.1 Signal's average power and PA behaviour

The Signal's average power affects the junction temperature and, consequently, the PA behaviour. To observe the effect of the signal's average power on the transmitter behaviour, the signal is transmitted at different IPBO values while the transmitter's signal quality and ACPR of the transmitted signal has been recorded. Figure 5.5 (a)-(b)-(c) shows the power spectral density of the transmitted signal at 0, 2 and 4 dB IPBO. The signal is linearized using PTI-DPD, AP-DPD, Maxp-DPD and reference [15]. Figure 5.6 presents the ACPR of the transmitted signal versus signal back-off power using different techniques. The results show that the PTI-DPD provides a close performance to reference [15] when operating at room temperature. The Maxp-DPD is extracted at maximum output power, and as expected, it shows its best linearization performance at the maximum output power, 0 dB IPBO. The AP-DPD presents a consistently poor performance across the power range. For further analysis, the measured AM/AM and AM/PM of the PA output

(a)

(b)

(c)
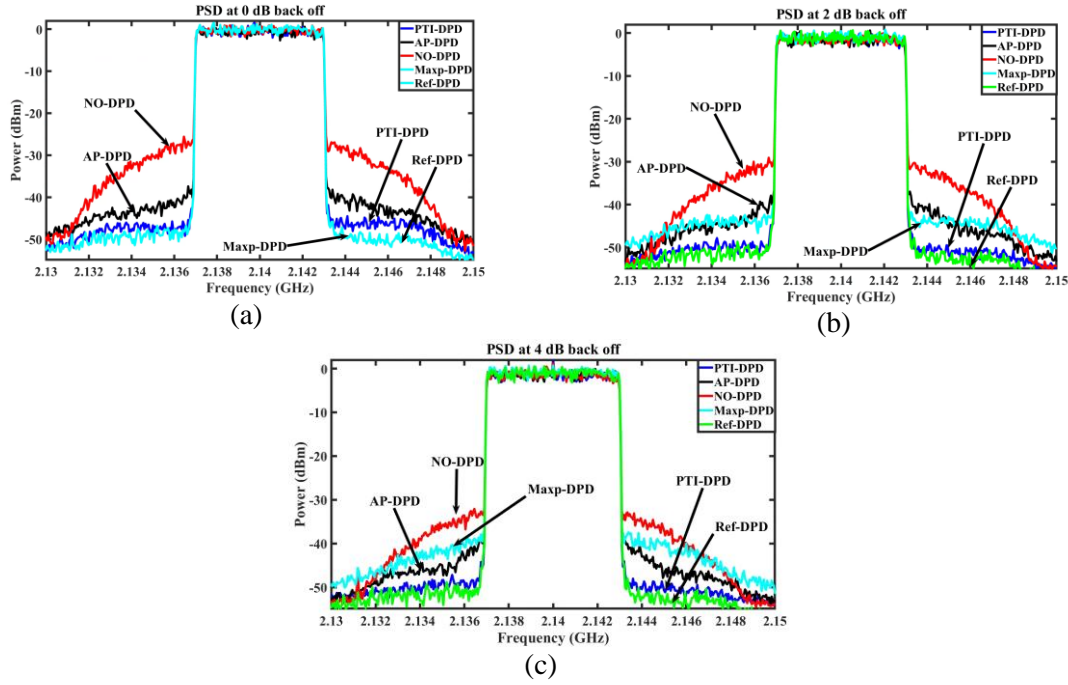
Figure 5.5 Average power dependency analysis of linearization methods in the term of (a) PSD at 0 dB, (b) PSD at 2 dB, (c) PSD at 4 dB IPBO [51]
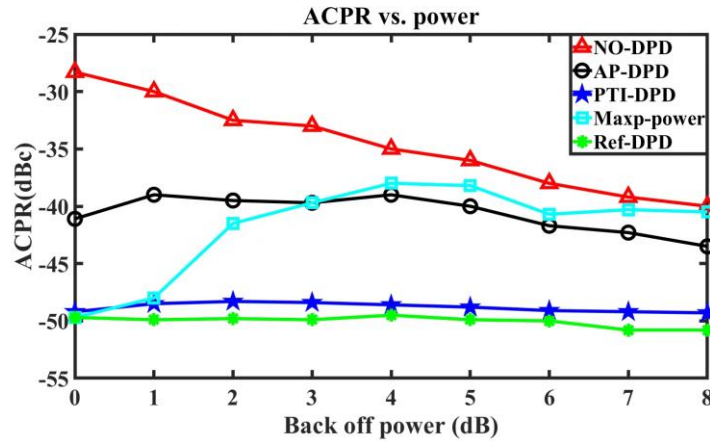
Figure 5.6 The ACPR of the transmitted signal at different power levels using different DPD techniques.

without DPD (no-DPD) and linearized using PTI-DPD, when driven at 0 dB IPBO, are presented

in Figure 5.7, respectively. The results show that the PTI-DPD is able to successfully compensate
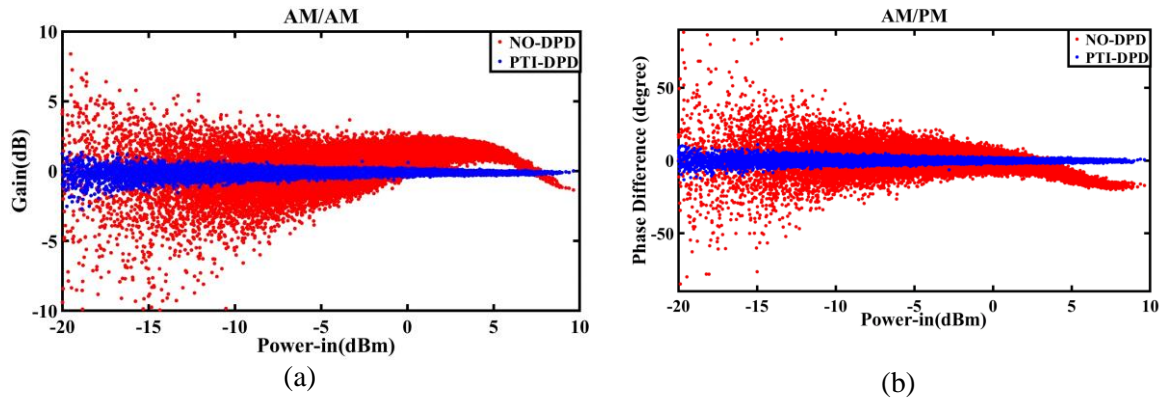
Figure 5.7 (a) AM/AM (b) AM/PM of the received signal using PTI-DPD [51]

for most of the nonlinearities and dispersions on the AM/AM and AM/PM of the transmitted signal

at 0 dB IPBO.

5.5.2 Measurement of the joint effect of base plate temperature and signal's average power variations on PA behaviour

To study the effect of base plate temperature on the PA's behaviour, the base plate temperature is swept from 27° to 69° Celsius while monitoring the PA's behaviour. In the first experiment, the signal is linearized at 0 IPBO and 27° Celsius using the Reference [15] and transmitted through the PA. In the next step, the base plate temperature is increased slowly while the output signal is monitored. The results reported in Figure 5.8 shows that the model of reference [15] is not able to compensate for the distortions added to the signal due to temperature variation. As a result, the model of reference [15] loses almost 8 dB of its performance at 69° Celsius. The same experiment is repeated for the signal linearized using PTI-DPD. The results show that the
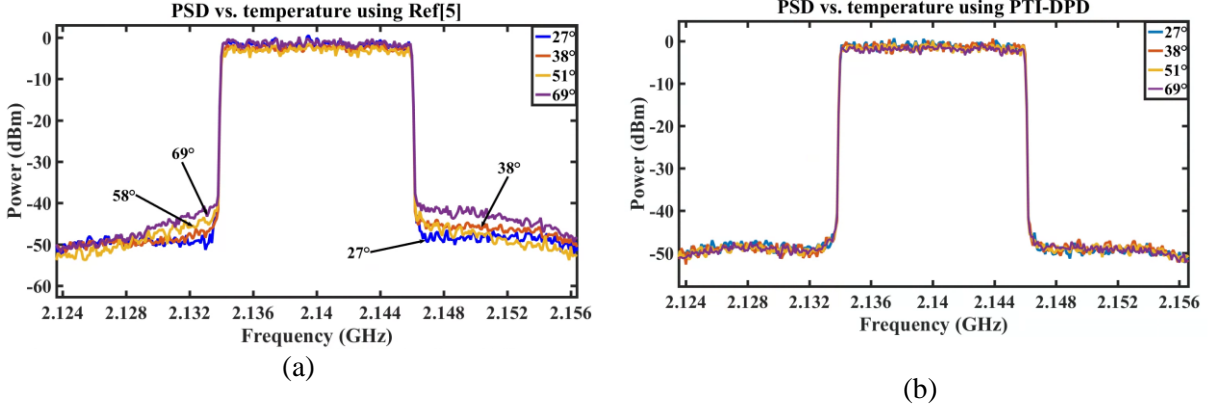
Figure 5.8 Temperature dependency analysis of the linearized transmitter (a): PSD at 0 dB IPBO using ref [15], (b): PSD at 0 dB IPBO using PTI-DPD [51].

PTI-DPD successfully compensates for any extra distortion added to the PA due to temperature variation.

To observe the effect of base plate temperature and signal's average power jointly, base plate temperature is swept from 27° to 69°, meanwhile, the signal's average power is varied from 0 to 8 dB IPBO with one 1 dB steps. Figure 5.9 shows the ACPR of the transmitted linearized using different DPD techniques signal under test conditions. The results show that the Maxp-DPD is not able to cope with temperature variation and it loses a part of its linearization performance due to temperature variation. Moreover, the effects of base plate temperature increment are more severe in higher power output signals. As the signal's average power levels down, the destructive effect of the base plate temperature on the PA's behaviour is reduced too. The same experiment is

ACPR-Max-DPD   ACPR-AP-DPD

(a)   (b)

ACPR-PTI-DPD   ACPR-Ref[5]

(c)   (d)
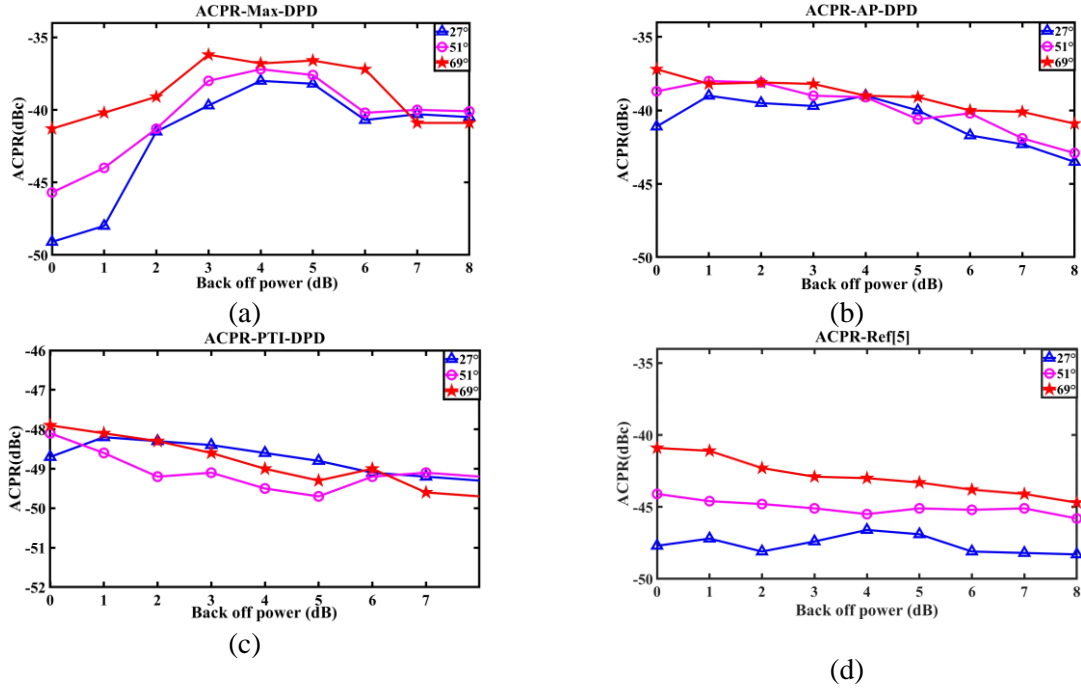
@ 2021 IEEE

Figure 5.9 Temperature dependency analysis of the linearized transmitter (a): ACPR of Maxp-DPD, (b) ACPR of AP-DPD, (c) ACPR of PTI-DPD, (d) ACPR of ref [15] across IPBO and base plate temperature variation [51].

repeated for AP-DPD, PTI-DPD and reference [15] while the PA behaviour is recorded. Figure 5.9 (b) shows the AP-DPD under the testing conditions. The results show that for a higher power signal with the highest base plate temperature, the AP-DPD presents the poorest linearization performance. Similarly, the performance of the DPD model in reference [15] degrades as the temperature increases and it is not able to adapt itself to temperature variation. On the other hand, the PTI-DPD monitors both base plate temperature and signal's average power and compensates for the extra distortions added to the system. Figure 5.9 (c) shows the transmitter's behaviour linearized using the PTI-DPD. The results show that, contrary to the other techniques, the PTI-DPD provides a stable linearization performance with maximum 2 dB fluctuation in the ACPR.

5.6　Conclusion

The radio base stations are supposed to operate in various geographical locations where the base plate temperature can range from -40° in North America to 50° in the Middle East. In addition, based on the users' locations, the transmitters are required to operate at different output power levels. The base plate temperature and signal's average power both influence the PA behaviour. The results showed that when not compensating for base plate temperature and signal's average power variation, the transmitted signal quality will substantially degrade. To address this issue, this chapter introduced PTI-DPD where the signal's average power, base plate temperature and modulated signal are the model input. The proposed PTI-DPD provides continuous linearization when the ambient temperature and signal's average power are varied.

6 Conclusion

An increasing need to design high-speed internet connections with limited available frequency spectrum has led wireless standards to use high-density modulations and multiple access techniques that generate signals with high peak-to-average power ratios along with MIMO and beamforming transmitter architectures. The high peak-to-average power ratios puts stringent requirements on the power consumption of the base station, which is a big design challenge that needs to be considered carefully. Indeed, power efficiency is a very important goal for any wireless system design as it reduces the design and maintenance cost and, in the case of battery-powered systems, increases the battery's lifetime. To achieve the power efficiency goal, the PAs are pushed to operate as close as possible to the saturation. Operating in this efficient but highly nonlinear region increases the distortions introduced to the signal. Moreover, MIMO and beamforming transmitters suffer from strong nonlinearities and crosstalk between the RF channels that are direction dependent. To compensate for these destructive effects of direction dependency of the distortions, crosstalk between the RF channels and PA nonlinearities, Digital Pre-Distortion (DPD) is used. In some applications, the wireless transmitter is supposed to operate at different environmental conditions and operating power. It has been observed that baseplate temperature and signal's average power can change the PA behaviour. Therefore, these parameters need to be considered in DPD design too.

The complexity of distortions results in complex mathematical equations for equation-based modeling. This makes it very computationally expensive to implement such models for practical applications. In contrast to equation-based modeling, artificial intelligence-based DPD can offer a solution with improved performance in scenarios where the number of modeling parameters are more than one. The main advantage of AI over conventional techniques is that it

can include various environmental and transmitter setting parameters without the need for a closed-form expression relating them.

In conclusion, the artificial intelligence based DPD provides a robust and adaptable linearization performance in different environmental and transmitter settings conditions. AI can help reduce the hardware complexity in massive MIMO beamforming transmitters where hundreds of RF channels are squeezed into a compact area. Moreover, the AI based DPDs eliminate the need for a dedicated observation path for each subarray and consequently reduce the cost and hardware complexity. The strong modeling capability of artificial neural networks and the advancement in customized AI processing units make it feasible to benefit from AI in DPD design for massive MIMO applications.

The AI is able to help designing more compact and accurate beamforming transmitters. AI can include various parameters into the DPD model making the transmitter performance more robust in different operating conditions. Contemporary to this work, the researchers have come up with a similar solution to DPD design. They have used AI to provide continuous linearization at different frequency bandwidth and output power. It is expected that more research groups will focus on this emerging area in DPD design. To come up with more innovative ways of using AI in DPD design more research need to be done in this area. One of the important areas is reducing the AI model identification complexity hardware implementation for commercial applications.

7   List of Publication

Publications:

A. Motaqi, M. Helaoui, N. Boulejfen, W. Chen and F. Ghannouchi, "Artificial Intelligence based Power-Temperature Inclusive Digital Pre-Distortion," in IEEE Transactions on Industrial2. Electronics, doi: 10.1109/TIE.2021.3128894.

A. Motaqi, M. Helaoui, A. Abdelaziz, W. Chen and F. M. Ghannouchi, "Averaged and Cluster DPDs for Beamforming Applications," *2021 European, Microwave Conference Proceeding*, London, April 2022.

Submitted:

A. Motaqi, M. Helaoui, A. Abdelaziz and F. M. Ghannouchi, "Hardware calibration and impairment mitigation in phased array antenna transmitter for enhanced beamforming linearization performance", 2022, Springer, Analog Integrated Circuits and Signal Processing

A. Motaqi, M. Helaoui, N. Boulejfen, W. Chen and F. Ghannouchi, "Angle Inclusive Digital Pre-Distorter for Beamforming Phased Array Transmitters Using Convolutional Neural Networks", IEEE Transactions on Microwave Theory and Technique TMTT.

Patents
filed:
A. Motaqi, F. Ghannouchi, M. Helaoui, "Apparatus and Method for Digital Predistortion for Active Massive Phased Array Systems", US Patent 2021

8 References

[1] S. Nagul, "A review on 5G modulation schemes and their comparisons for future wireless communications," 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), 2018, pp. 72-76, doi: 10.1109/SPACES.2018.8316319.

[2] G. R. Nikandish, R. B. Staszewski and A. Zhu, "Unbalanced Power Amplifier: An Architecture for Broadband Back-Off Efficiency Enhancement," in IEEE Journal of Solid-State Circuits, vol. 56, no. 2, pp. 367-381, Feb. 2021, doi: 10.1109/JSSC.2020.3014244.

[3] G. Baudoin and P. Jardin, "Adaptive polynomial pre-distortion for linearization of power amplifiers in wireless communications and WLAN," EUROCON'2001. International Conference on Trends in Communications. Technical Program, Proceedings (Cat. No.01EX439), 2001, pp. 157-160 vol.1, doi: 10.1109/EURCON.2001.937787.

[4] K. J. Muhonen, M. Kavehrad and R. Krishnamoorthy, "Look-up table techniques for adaptive digital predistortion: a development and comparison," in IEEE Transactions on Vehicular Technology, vol. 49, no. 5, pp. 1995-2002, Sept. 2000, doi: 10.1109/25.892601.

[5] 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, doi: 10.1109/78.552219.

[6] Lei Ding et al., "Memory polynomial predistorter based on the indirect learning architecture," Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE, 2002, pp. 967-971 vol.1, doi: 10.1109/GLOCOM.2002.1188221.

[7] 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 Trans. Signal Process.*, vol. 54, no. 10, pp. 3852-3860, Oct. 2006, doi: 10.1109/TSP.2006.879264.

[8] Taijun Liu, S. Boumaiza and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Trans.s Microw. Theory Techn.*, vol. 52, no. 3, pp. 1025-1033, March 2004, doi: 10.1109/TMTT.2004.823583.

[9] R. Hongyo, Y. Egashira, T. M. Hone and K. Yamaguchi, "Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers," *IEEE Microwave and Wireless Components Lett.*, vol. 29, no. 2, pp. 146-148, Feb. 2019, doi: 10.1109/LMWC.2018.2888955.

[10] Y. Khawam, O. Hammi, L. Albasha and H. Mir, "Behavioral Modeling of GaN Doherty Power Amplifiers Using Memoryless Polar Domain Functions and Deep Neural Networks," in IEEE Access, vol. 8, pp. 202707-202715, 2020, doi: 10.1109/ACCESS.2020.3036186.

[11] X. Hu et al., "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers," *IEEE Trans.s Neural Networks and Learning Systems*, pp. 1-15, Feb.2021, doi: 10.1109/TNNLS.2021.3054867.

[12]H. Yin, Z. Jiang, X. -W. Zhu and C. Yu, "Over-the-air Behavioral Modeling of Millimeter Wave Beamforming Transmitters with Concurrent Dynamic Configurations Utilizing Heterogenous Neural Network," 2020 IEEE/MTT-S International Microwave Symposium (IMS), 2020, pp. 397-400, doi: 10.1109/IMS30576.2020.9223873.

[13] S. K. Dhar, A. Abdelhafiz, M. Aziz, M. Helaoui and F. M. Ghannouchi, "A Reflection-Aware Unified Modeling and Linearization Approach for Power Amplifier Under Mismatch and Mutual

Coupling," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 9, pp. 4147-4157, Sept. 2018, doi: 10.1109/TMTT.2018.2859959.

[14] S. A. Bassam, M. Helaoui and F. M. Ghannouchi, "Crossover Digital Predistorter for the Compensation of Crosstalk and Nonlinearity in MIMO Transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1119-1128, May 2009, doi: 10.1109/TMTT.2009.2017258.

[15] Y. Guo, C. Yu and A. Zhu, "Power Adaptive Digital Predistortion for Wideband RF Power Amplifiers With Dynamic Power Transmission," in IEEE Transactions on Microwave Theory and Techniques, vol. 63, no. 11, pp. 3595-3607, Nov. 2015, doi: 10.1109/TMTT.2015.2480739.

[16] O. Hammi, A. Kwan and F. M. Ghannouchi, "Bandwidth and Power Scalable Digital Predistorter for Compensating Dynamic Distortions in RF Power Amplifiers," in IEEE Transactions on Broadcasting, vol. 59, no. 3, pp. 520-527, Sept. 2013, doi: 10.1109/TBC.2013.2262692.

[17] Y. Yu, J. Cai, X. -W. Zhu, P. Chen and C. Yu, "Self-Sensing Digital Predistortion of RF Power Amplifiers for 6G Intelligent Radio," in IEEE Microwave and Wireless Components Letters, doi: 10.1109/LMWC.2021.3139018.

[18] P. K. Bailleul, "A New Era in Elemental Digital Beamforming for Spaceborne Communications Phased Arrays," in Proceedings of the IEEE, vol. 104, no. 3, pp. 623-632, March 2016, doi: 10.1109/JPROC.2015.2511661.

[19] D. H. Roper, W. E. Babiec, and D. D. Hannan, "WGS phased arrays support next generation DOD SATCOM capability," in Proc. IEEE Int. Symp. Phased Array Syst. Technol., 2003, pp. 82–87.

[20] V. Venkateswaran and A. van der Veen, "Analog Beamforming in MIMO Communications With Phase Shift Networks and Online Channel Estimation," in IEEE Transactions on Signal Processing, vol. 58, no. 8, pp. 4131-4143, Aug. 2010, doi: 10.1109/TSP.2010.2048321.

[21] Y. Karacora, A. Kariminezhad and A. Sezgin, "Hybrid Beamforming: Where Should the Analog Power Amplifiers Be Placed?," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 4719-4723, doi: 10.1109/ICASSP.2019.8683189.

[22] N. Peccarelli and C. Fulton, "A Mutual Coupling Approach to Digital Pre-Distortion and Nonlinear Equalization Calibration for Digital Arrays," *in 2019 IEEE Int. Symp. Phased Array Syst. Technol. (PAST)*, 2019, pp. 1-8, doi: 10.1109/PAST43306.2019.9020812.

[23] Available online at : https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-papers/3gpp-release-16-shifting-gears-to-increase-5g-speeds-on-multiple-network-highways/Samsung-3GPP-Release-16-Whitepaper.pdf

[24] X. Liu et al., "Beam-Oriented Digital Predistortion for 5G Massive MIMO Hybrid Beamforming Transmitters," in IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 7, pp. 3419-3432, July 2018, doi: 10.1109/TMTT.2018.2830772.

[25] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi and Z. Feng, "Linearization for Hybrid Beamforming Array Utilizing Embedded Over-the-Air Diversity Feedbacks," in IEEE Transactions on Microwave Theory and Techniques, vol. 67, no. 12, pp. 5235-5248, Dec. 2019, doi: 10.1109/TMTT.2019.2944821.

[26] X. Wang, Y. Li, C. Yu, W. Hong and A. Zhu, "Digital Predistortion of 5G Massive MIMO Wireless Transmitters Based on Indirect Identification of Power Amplifier Behavior With OTA Tests," in IEEE Transactions on Microwave Theory and Techniques, vol. 68, no. 1, pp. 316-328, Jan. 2020, doi: 10.1109/TMTT.2019.2944828

[27] C. Yu et al., "Full-Angle Digital Predistortion of 5G Millimeter-Wave Massive MIMO Transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2847-2860, July 2019, doi: 10.1109/TMTT.2019.2918450.

[28] Brihuega, L. Anttila and M. Valkama, "Neural-Network-Based Digital Predistortion for Active Antenna Arrays Under Load Modulation," *IEEE Microw. Wireless Components Lett.*, vol. 30, no. 8, pp. 843-846, Aug. 2020, doi: 10.1109/LMWC.2020.3004003.

[29] P. GRÖSCHEL ET AL., "A SYSTEM CONCEPT FOR ONLINE CALIBRATION OF MASSIVE MIMO TRANSCEIVER ARRAYS FOR COMMUNICATION AND LOCALIZATION," *IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES,* VOL. 65, NO. 5, PP. 1735-1750, MAY 2017, DOI: 10.1109/TMTT.2017.2663404.

[30] H. WEI, D. WANG, H. ZHU, J. WANG, S. SUN AND X. YOU, "MUTUAL COUPLING CALIBRATION FOR MULTIUSER MASSIVE MIMO SYSTEMS," *IN IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, VOL. 15, NO. 1, PP. 606-619, JAN. 2016, DOI: 10.1109/TWC.2015.2476467.

[31] L. J. FOGED, L. SCIALACQUA, F. SACCARDI, N. GROSS AND A. SCANNAVINI, "OVER THE AIR CALIBRATION OF MASSIVE MIMO TDD ARRAYS FOR 5G APPLICATIONS," *2017 IEEE INTERNATIONAL SYMPOSIUM ON ANTENNAS AND PROPAGATION & USNC/URSI NATIONAL RADIO SCIENCE MEETING, SAN DIEGO, CA,* 2017, PP. 1423-1424, DOI: 10.1109/APUSNCURSINRSM.2017.8072754.

[32] available online at: https://www.analog.com/en/products/ad9361.html

[33] available online at: https://www.xilinx.com/products/silicon-devices/soc/zynq-7000.html

[34] available online at : https://www.xilinx.com/support/documentation/sw_manuals/xilinx14_7/sysgen_ref.pdf

[35] available online at : https://www.putty.org/

[36] N. Zhang, H. Yin and W. Wang, "Hybrid beamforming for millimetre wave massive MU-MIMO systems with IQ imbalance," *IET Communications*, vol. 13, no. 6, pp. 776-785, 2 4 2019, doi: 10.1049/iet-com.2018.5661.

[37] M. Buchholz, A. Schuchert and R. Hasholzner, "Effects of tuner IQ imbalance on multicarrier-modulation systems," Proceedings of the 2000 Third IEEE International Caracas Conference on Devices, Circuits and Systems (Cat. No.00TH8474), 2000, pp. T65/1-T65/6, doi: 10.1109/ICCDCS.2000.869880.

[38] D. PARKER AND D. C. ZIMMERMANN, "PHASED ARRAYS - PART 1: THEORY AND ARCHITECTURES," *IN IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES*, VOL. 50, NO. 3, PP. 678-687, MARCH 2002, DOI: 10.1109/22.989953.

[39] Xinying Zhang, A. F. Molisch and Sun-Yuan Kung, "Variable-phase-shift-based RF-baseband codesign for MIMO antenna selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4091-4103, Nov. 2005, doi: 10.1109/TSP.2005.857024.

[40] M. Abualhayja'a and M. Hussein, "Comparative Study of Adaptive Beamforming Algorithms for Smart antenna Applications," 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), 2021, pp. 1-5, doi: 10.1109/ICCSPA49915.2021.9385725.

[41] M. K. Baburaj, S. K. Ambat and V. S. Sheeba, "Compressive beamforming using greedy algorithms," 2016 International Conference on Communication Systems and Networks (ComNet), 2016, pp. 81-86, doi: 10.1109/CSN.2016.7823991.

[42] https://www.xilinx.com/products/design-tools/vitis/vitis-ai.html

[43]L. Shao, F. Zhu and X. Li, "Transfer Learning for Visual   Categorization: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 1019-1034, May 2015, doi: 10.1109/TNNLS.2014.2330900.

[44] O. Abdel-Hamid, A. Mohamed, H. Jiang and G. Penn, "Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition," 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012, pp. 4277-4280, doi: 10.1109/ICASSP.2012.6288864.

[45] E. Ng, Y. Beltagy, G. Scarlato, A. Ben Ayed, P. Mitran and S. Boumaiza, "Digital Predistortion of Millimeter-Wave RF Beamforming Arrays Using Low Number of Steering Angle-Dependent Coefficient Sets," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 11, pp. 4479-4492, Nov. 2019, doi: 10.1109/TMTT.2019.2924893.

[46] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions* (Applied Mathematics Series 55), 1 st ed. Washington, DC, USA: NBS, 1964, pp. 355.

[47] Poletaev, I. & Pervunin, Konstantin & Tokarev, M.. (2016). Artificial neural network for bubbles pattern recognition on the images. Journal of Physics Conference Series. 754. (072002)-13. 10.1088/1742-6596/754/7/072002.

[48] S. Boumaiza and F. M. Ghannouchi, "Thermal memory effects modeling and compensation in RF power amplifiers and predistortion linearizers," in IEEE Transactions on Microwave Theory and Techniques, vol. 51, no. 12, pp. 2427-2433, Dec. 2003, doi: 10.1109/TMTT.2003.820157.

[49] S. K. Dhar, M. Helaoui and F. M. Ghannouchi, "Temperature Dependent Robust Behavioral Modeling of Non-Linear Power Amplifier," 2018 Asia-Pacific Microwave Conference (APMC), 2018, pp. 378-380, doi: 10.23919/APMC.2018.8617427.

[50] W. Wei, O. K. Jensen and J. H. Mikkelsen, "Self-heating and memory effects in RF power amplifiers explained through electro-thermal," 2013 NORCHIP, 2013, pp. 1-4, doi: 10.1109/NORCHIP.2013.6702015.

[51] A. Motaqi, M. Helaoui, N. Boulejfen, W. Chen and F. M. Ghannouchi, "Artificial Intelligence-Based Power-Temperature Inclusive Digital Predistortion," in IEEE Transactions on Industrial Electronics, vol. 69, no. 12, pp. 13872-13880, Dec. 2022, doi: 10.1109/TIE.2021.3128894.