Digital Predistortion of Millimeter-Wave Phased Antenna Arrays

by

Eric Ng

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

Waterloo, Ontario, Canada, 2018

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

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

Abstract

The cost of deployment of reliable, high-throughput, fifth-generation (5G) millimeter-wave (mm-wave) base stations will depend significantly on the maximum equivalent isotropically radiated power (EIRP) that the base stations can transmit. High EIRP can be generated using active beamforming antenna arrays with large apertures and driven by an array of power amplifiers (PAs). However, given the tight half-wavelength lattice that the arrays must retain to ensure a wide beam steering range, the achievable EIRP quickly becomes thermally-limited. Efficient power amplification is thus imperative to low-cost and reliable beamforming antenna array design. This work considers the application of Digital Predistortion (DPD) as an efficiency-enhancement technique for mm-wave beamforming antenna arrays. Two RF beamforming configurations were considered and corresponding DPD schemes were investigated.

First, a single-input single-output (SISO) DPD is proposed that can linearize a single-user RF beamforming array in the presence of non-idealities such as PA load modulation and variation of phase shifter gain with phase. The SISO DPD relies on a feedback signal which reflects a coherent summation of the PA outputs. The SISO DPD is then validated by measurement of a 4-element and 64-element array at 28 GHz with 800 MHz modulated signals using a single over-the-air feedback signal. The SISO DPD uses different sets of coefficients to cope with changes in both linear and non-linear distortions as the beam is steered.

Second, the SISO DPD formulation is extended to multi-user RF beamforming to linearize multiple sub-arrays. In this configuration, non-negligible inter-user interference can affect the DPD training. To address the interference, a linear estimate of the interference is calculated and canceled for each user before extracting the SISO DPD coefficients in each sub-array. The SISO DPD with interference cancellation is validated by measurement of a 2-user 2x64-element subarray hybrid at 28 GHz with 800 MHz modulated signals across different combinations of steering angles for the two users.

Acknowledgements

This thesis would not have been possible without the helpful guidance of my supervisor and professors and support from fellow colleagues at the university. It was a pleasure to have worked with and learned from this community, and I thank them for their patience and care.

I would also like to personally thank everyone that has been or currently are a mentor, teacher, co-worker, family member or friend of mine during my study. I have had countless positive experiences during my education and I am truly grateful for all the opportunities I have had.

Special thanks are extended to members from Anokiwave and Keysight Technologies for providing the equipment used in the thesis.

Dedication

This is dedicated to Jack, Jennifer, and Michael for always being there for me.

Table of Contents

List of Figures

List of Abbreviations

5G Fifth-Generation 7

ACF Aperture Consumption Factor 9

ACPR Adjacent Channel Power Ratio 37

AF Array Factor 5

AM Amplitude Modulation 12

ANN Artificial Neural Network 14

AWG Arbitrary Waveform Generator 36

BTS Base Tranceiver Station 2

CAGR Compound Annual Growth Rate 1

CCDF Complementary Cumulative Distribution Function 11

CW Continuous Wave 5

DAC Digital-to-Analog Converter 52

DC Direct Current 8

DFT Discrete Fourier Transform 47

DPD Digital Predistortion 12

EIRP Effective Isotropic Radiated Power 4

EMBB Enhanced Mobile Broadband 1

EVM Error Vector Magnitude 47

FCC Federal Communications Commission 3

FSPL Free-Space Path Loss 3

GMP Generalized Memory Polynomial 14

IB In-band 38

IC Integrated Circuit 42

IF Intermediate Frequency 65

IFFT Inverse Fast Fourier Transform 11

IoT Internet-of-Things 1

IQ In-phase and Quadrature 36

ITU International Telecommunications Union 2

LO Local Oscillator 27

LOS Line-of-Sight 22

MEC Mobile Edge Computing 1

MIMO Multiple-Input Multiple-Output 55

MISO Multiple-Input Single-Output 61

mm-wave Millimeter-Wave 3

MMC Massive Machine-Type Communication 1

MRT Maximum Ratio Transmission 23

NMSE Normalized Mean-Squared Error 37, 38

OFDM Orthogonal Frequency-Division Multiplexing 10

OOB Out-of-band 23

OTA Over-the-air 38

PA Power Amplifier 8

PAPR Peak-to-Average Power Ratio 11

PM Phase Modulation 12

QAM Quadrature Amplitude Modulation 10

QPSK Quadrature Phase-Shift Keying 10

RF Radio Frequency 8

SISO Single-Input Single-Output 13

SNR Signal-to-Noise Ratio 3

TOR Transmission Observation Receiver 13

TRP Total Radiated Power 8

ULCC Ultra-Reliable Low-Latency Communication 1

VNI Visual Networking Index 1

WLAN Wireless Local Area Networks 51

Chapter 1

Introduction

1.1 The Need for Higher Data Capacity

Since the introduction of the first Internet networks in the 1970s, the requirements for expeditious information transfer have been growing rapidly year after year. Many have seen the impact of the development of the transistor or its commercial manifestation in Moore's Law in the changes to their day-to-day life; from the reliance of personal computers at home or work, to the enjoyment garnered from digitally-rendered movies and video games. While Moore's Law describes the trend experienced between 1970 to 2010 which saw the density of transistors increasing by roughly 40% each year (doubling every 2 years), a similar phenomenon is already taking shape in the form of Internet data traffic. In particular, Cisco's Visual Networking Index (VNI) predicts a Compound Annual Growth Rate (CAGR) of 46% for the global demand for mobile data between 2016 to 2021 [7]; double the rate of fixed data growth and rivaling that of Moore's Law. This rapid rate of growth is set to bring forth immense benefits and challenges in the area of communications for at least the next decade.

The data growth is anticipated to stem from three primary usage scenarios [8]. These scenarios include Enhanced Mobile Broadband (EMBB), Ultra-Reliable Low-Latency Communication (ULCC), and Massive Machine-Type Communication (MMC), as illustrated in Fig. 1.1. There has already been much research and commercial interest to realize these usage scenarios and associated applications. EMBB applications include high-definition video streaming in both urban and rural environments [9]. ULLC scenarios involve the use of Mobile Edge Computing (MEC) [10], which include driverless cars [11], remote medicine [12], and augmented/virtual reality [13]. Finally, MMC builds upon Internet-of-Things (IoT)

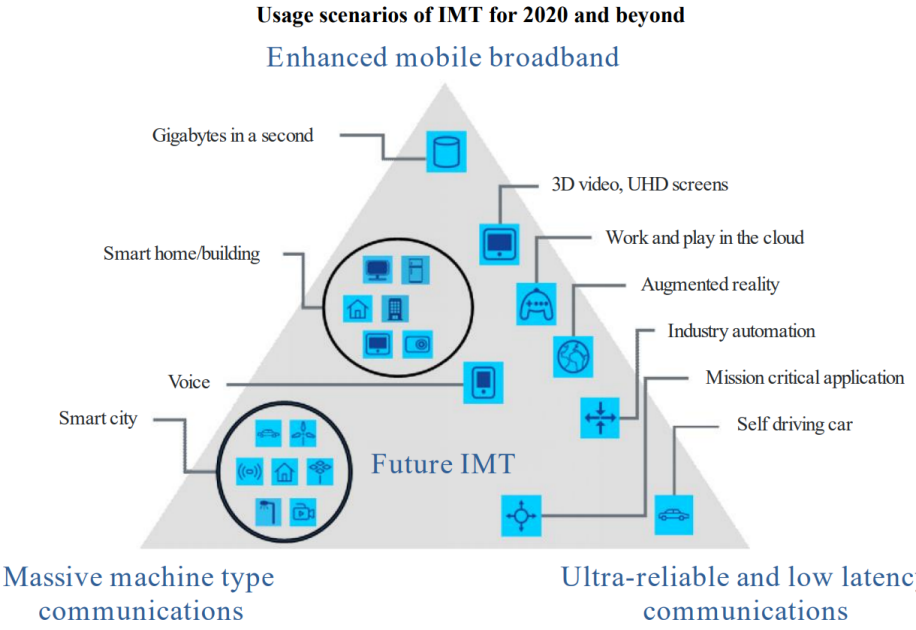

Figure 1.1: Usage scenarios as envisioned by the International Telecommunications Union in 2020 and beyond.

philosophy [14], which promises smart cities and sensor-assisted manufacturing and agriculture.

To successfully realize the many applications of next-generation mobile networks, higher aggregate data capacity is required. The aggregate demanded area data capacity, C_d (in bits per second per area), can be written as the product of the number of connections per area, N_c, and the average data-rate per connection, C_u. This demanded data must be significantly less than the available data capacity from the network per unit area, C_a, which is comprised of the area spectral efficiency (in bits per second per hertz per Base Tranceiver Station (BTS)), η_B, the occupied spectrum bandwidth, B, and the density of base-stations D_b. The relationship is written as:

$$C_\mathrm{d} = N_c C_u \ll C_\mathrm{a} = \eta_B B D_b. \tag{1.1}$$

It is evident from the above equation that to increase C_a, one must increase η_B, B, and/or D_b. The International Telecommunications Union (ITU) predicts that at least a 100 times increase in C_a [8] is necessary to fulfill the requirements of the next-generation standards. There are many methods to arrive at the 100-factor increase. However, the combination

with the lowest cost of implementation is likely to be favored by the mobile network operators in practical deployments. This restricts significant increases in D_b, as the cost of deployment would heavily depend on the required number of BTS per unit area.

1.2 The Case for Millimeter-Wave

One method that can be used to help achieve the necessary hundred-fold increase in capacity is the bandwidth, B, also viewed as the rate at which information is sent across the channel. From the Shannon-Hartley theorem, it is known that given a constant and reasonably high Signal-to-Noise Ratio (SNR) ratio, the capacity will increase proportional to the bandwidth of the signal. The increased bandwidth requires the network operator to have access to more spectrum, which, at its current state, is limited. Current generation cellular networks in the US occupy not more than 780 MHz of spectrum total [1] in bands that extend up to 3 GHz, as shown in Fig. 1.2. The spectrum up to 3 GHz is heavily congested but highly valued; the most recent major spectrum auction held in the United States saw 62 MHz of bandwidth total around the 700 MHz band sold for over $19 billion USD [15]. Fortunately, more bandwidth can be found upwards of 3 GHz; at frequencies between 30 to 300 GHz known as Millimeter-Wave (mm-wave). In this range, an abundance of spectrum has become recently available, with the Federal Communications Commission (FCC) opening up 10.85 GHz (about 14 times the currently used spectrum) for auction [2] and an additional 26 GHz being studied for future use by the ITU [3].

1.3 Challenges at Millimeter-Wave

Friis' formula for the power at the receiver (P_{rx}) in ideal free-space propagation can be expressed in terms of the gain of the transmitter antenna (G_{tx}), gain of the receiver antenna (G_{rx}), transmitted power (P_{tx}), and the Free-Space Path Loss (FSPL) can be expressed as:

$$\mathrm{P}_{rx}(\mathrm{dBm}) = \mathrm{G}_{tx}(\mathrm{dBi}) + \mathrm{P}_{tx}(\mathrm{dBm}) + \mathrm{G}_{rx}(\mathrm{dBi}) + \mathrm{FSPL}, \tag{1.2}$$

where

$$\mathrm{FSPL}(\lambda_0)(\mathrm{dB}) = 10\log_{10}(\frac{\lambda_0}{4\pi R})^2, \tag{1.3}$$

R is the distance between the transmitter and receiver, and λ_0 is the wavelength corresponding to the frequency of operation, f, in a free-space vacuum ($\lambda_0(f) = \frac{c_0}{f}$).

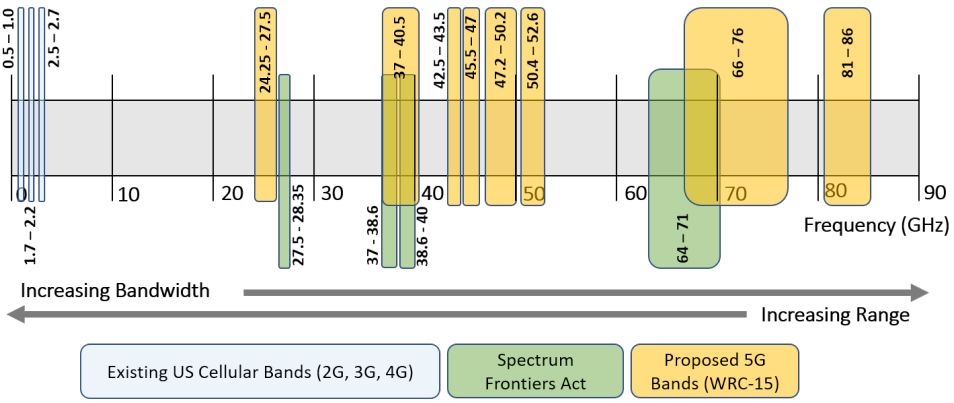

Figure 1.2: Allocation of spectrum in 2G, 3G and 4G in United States [1] compared to the mm-wave spectrum to be auctioned as part of the FCC's Spectrum Frontiers Act [2] and the proposed mm-wave bands from the World Radiocommunications Conference, 2015 [3].

As shown from the above equations, the losses in free-space increases with the square of the ratio of $\frac{\lambda_0}{R}$. In other words, with all else being constant, since λ_0 at mm-wave is at least 10 times smaller than λ_0 at 3 GHz, the coverage area to maintain the same received power would decrease by a factor of 100 when going from 3 GHz to 30 GHz. Compounding with this issue is that atmospheric attenuation tends to increase with frequency, as indicated in Fig. 1.3. This effect would then severely limit the coverage area of mm-wave BTS, leading to high cost of deployment.

1.4 The Phased Array Solution

To alleviate the coverage area challenge at mm-wave, one can attempt to increase the Effective Isotropic Radiated Power (EIRP) at the BTS, defined as:

$$\mathrm{EIRP} = \mathrm{G}_{tx}(\mathrm{dBi}) + \mathrm{P}_{tx}(\mathrm{dBm}). \tag{1.4}$$

Friis' formula can then be rewritten in terms of EIRP as:

$$\mathrm{P}_{rx}(\mathrm{dBm}) = \mathrm{EIRP}(\mathrm{dBm}) + \mathrm{G}_{rx}(\mathrm{dBi}) + \mathrm{FSPL}(\lambda_0(f)). \tag{1.5}$$

To maintain the same P_{rx} at the same distance for 30 GHz compared to 3 GHz, at least a 20 dB increase in EIRP is needed to compensate for the FSPL alone.

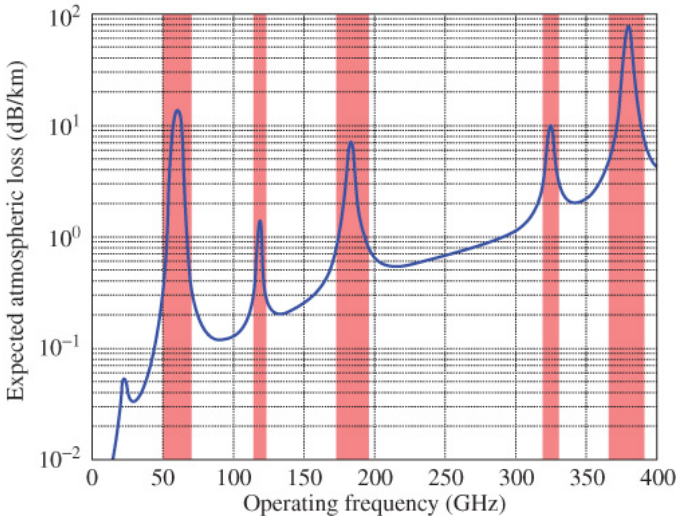

Figure 1.3: Atmospheric attenuation at mm-wave [1], the highlighted bars indicate areas of high absorption.

Now to increase the EIRP, one must either increase G_{tx} or P_{tx}. One method of increasing G_{tx} is by increasing the aperture of the antenna through use of phased/coherent combining of multiple antenna elements. This is illustrated for a linear (1D) array of isotropic elements spaced d apart in Fig. 1.4, where r_n shows the direction of propagation of the n'th antenna towards an infinitely-far receiver at θ. As shown in Fig. 1.4, when the receiver is sufficiently far away from the transmitter (i.e. in the far-field such that $R > \frac{2D^2}{\lambda_0}$ where D is the largest dimension of the array), the distance between the paths that each antenna takes is $d\cos(\theta)$. Assuming the $n \in 1, 2, ..., N$ antennas are excited by a Continuous Wave (CW) sinusoid with coefficient α_n (e.g. $i(t) = \alpha_n \cos(\omega t)$, where $i(t)$ is the instantaneous current of the excitation) the combined coefficient in the far-field, or Array Factor (AF) is expressed as:

$$\mathrm{AF} = \sum_{n=1}^{N} \alpha_n e^{j(n-1)\psi}, \tag{1.6}$$

where $\psi = kd\cos(\theta)$ and $k = \frac{2\pi}{\lambda_0(f)}$ is the wave number. If the α_n are identical but each phased by $-\psi$ apart, i.e. $\alpha_n = \alpha e^{-j(n-1)\psi}$, then the elements combine in phase so that $\mathrm{AF} = N\alpha$, which is N times larger than the single isotropic antenna. Subsequently, with proper phasing of the antenna excitation, it is possible to express the antenna gain for the array as $G_{tx}(\mathrm{dBi}) = G_a(\mathrm{dBi}) + 20\log_{10}(N)$, where G_a is the gain of a single antenna above

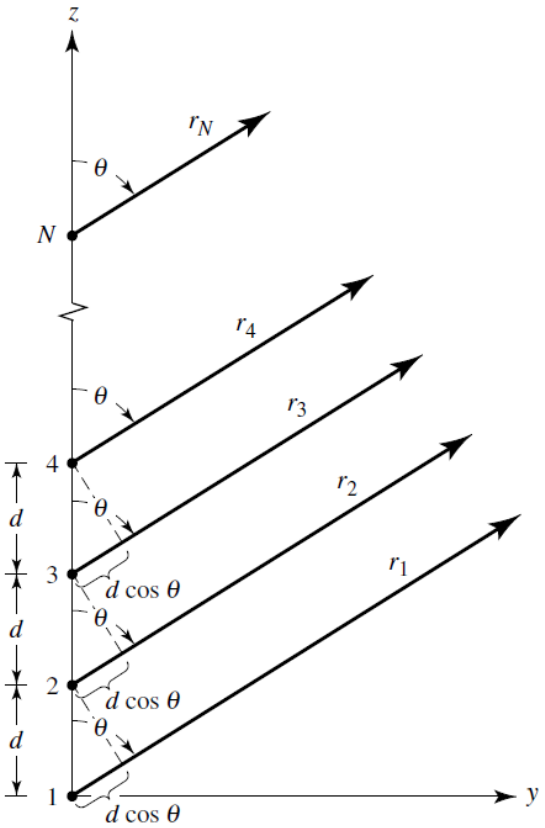

Figure 1.4: Far-field geometry of an N-element linear array of isotropic sources along the z-axis [4].

an isotropic element. Such arrays are called beamforming, or beamsteering arrays. The 20 dB difference between the 30 GHz and 3 GHz FSPL can then be compensated using an array size $N = 10$.

It is interesting to note that if $d = \frac{\lambda_0(f)}{2}$ (as is required to achieve good beamsteering range), a single half-wavelength antenna designed at $f_l = \frac{f}{N}$ would have the same physical size as an array of N half-wavelength antennas designed at f. This highlights how arrays synthesized at mm-wave will not need to be much larger than the single antennas used in current 3G and 4G systems.

The phased array technique is not limited to 1D arrays, and more commonly, the antenna elements can be tiled with spacing d_x and d_y as a 2D/planar array shown in Fig. 1.5. In this configuration, the corresponding AF is:

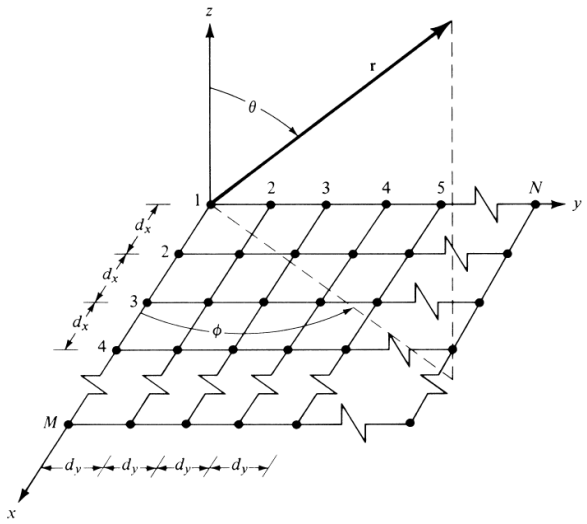

Figure 1.5: Geometry of a 2D planar array [4] with a total of $M \times N$ elements.

$$\mathrm{AF} = \sum_{n_x=1}^{M} \sum_{n_y=1}^{N} \alpha_{m,n} e^{jk\sin(\theta)((n_x-1)d_x\cos(\phi)+(n_y-1)d_y\sin(\phi))}. \tag{1.7}$$

These arrays can be used to steer the direction of their beam (direction of maximum AF) by selection of the excitation phase at each element. Multiple arrays can also be combined to steer to multiple directions. The magnitude of the AF for an 8×8 planar isotropic array when plotted over $\theta \in [0°, 90°]$ and $\phi \in [0°, 360°]$ is shown in Fig. 1.6.

It is evident from Fig. 1.6 that the power from a phased array is spatially-directed, and this spatial-directivity can lead to sufficient decorrelation between the signals received from multiple arrays such that multiple independent streams can be transmitted and received simultaneously. Since the bandwidth has not changed for each beam, this feature, called spatial multiplexing, allows the spectral efficiency, η_B, in multi-user beamforming arrays to increase. Recalling from $C_\mathrm{a} = \eta_B B D_b$, it is the combination of the increase in B and η_B without increasing D_b that makes mm-wave phased arrays such an attractive solution for Fifth-Generation (5G) transmitters.

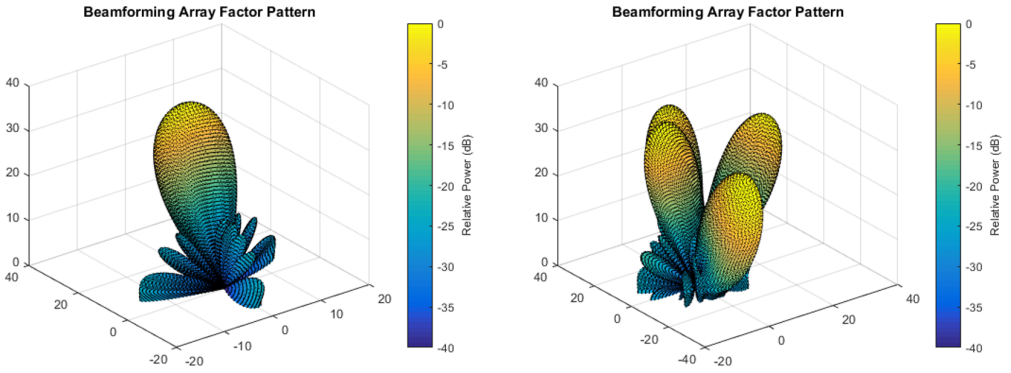

Figure 1.6: AF of a single 8×8 array (left) and 4 arrays (right), each phased to steer their beams to different angular directions.

1.5 Implementation Challenges for Practical Millimeter-Wave Phased Arrays

There are many methods of implementing phased arrays. One of them is the active Radio Frequency (RF) beamforming array, shown in Fig. 1.7. In this configuration, the gain and phase shifts are performed after upconversion but before the Power Amplifier (PA). As opposed to passive phased arrays where the amplifier is placed before the gain and phase shifts and the divider, the PAs directly drive the antenna and hence this implementation improves the linearity (highest linear output power) of the transmitter. All else being equal, this configuration will allow the highest EIRP to be achieved for a given amount of Direct Current (DC) power draw. The challenge in this implementation is that since there are as many PAs as there are antennas, the PAs must be tightly integrated and so large arrays must be able to effectively dissipate the heat. The tight area constraints are imposed due to the $d \leq \frac{\lambda_0}{2}$ grid that the antennas must be placed in to ensure no grating lobes as the beam is steered far from the array broadside ($\theta = 0°$) [4]. If the heat is not dissipated effectively enough, the high operating temperatures will accelerate the aging rate of the devices, leading to shorter basestation lifetime and reliability issues. Active cooling, such as by use of fans, are also not desirable as they consist of moving parts and are prone to failure, leading to higher maintenance costs as well.

It is due to the increasing heat dissipation requirements that the array's EIRP becomes thermally-limited. At the time of writing, state of the art commercial arrays, such as in [16,17], do not approach close to the maximum Total Radiated Power (TRP) allowed by the

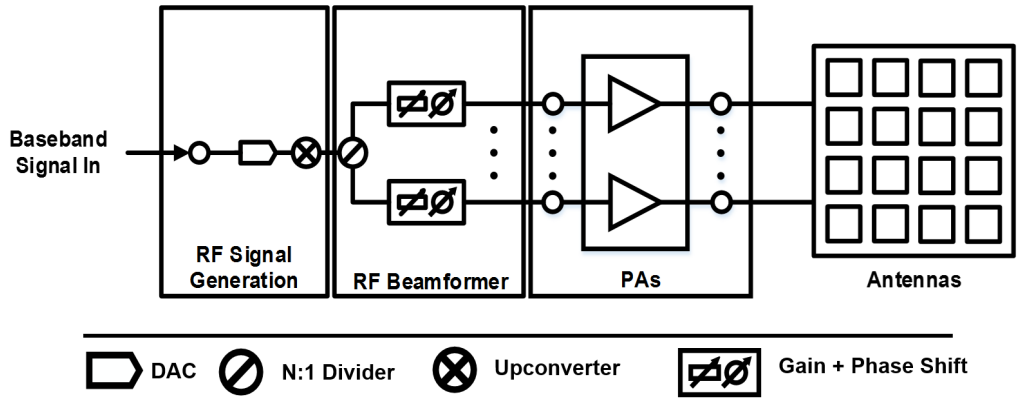

Figure 1.7: Active RF beamforming array, where the gain and phase settings for each antenna are performed after upconversion but before the PA.

5G standards body [18]. It is perhaps useful to have a figure-of-merit that quantifies phased array designs that will fare better when dealing with the thermal limitation. The Aperture Consumption Factor (ACF), can be defined using EIRP and the DC power consumption of the phased array RF front-end, $\mathrm{P_{DC}}$, as:

$$\mathrm{ACF} = \frac{\mathrm{EIRP}}{\mathrm{P_{DC}}} = \frac{\mathrm{G}_t \mathrm{P}_t}{\mathrm{P_{DC}}}. \tag{1.8}$$

As evident in the definition, the ACF can be improved by either minimizing the DC power consumption or increasing the EIRP. Note that while some definitions of $\mathrm{P_{DC}}$ include the power consumption from baseband signal generation, the mixed-signal circuitry is typically not tightly integrated with the phased array front-end and will not have as significant an impact on the thermal limitation as the front-end components.

One can imagine many methods of increasing the ACF. For example, increasing the number of antenna elements in the array is a possibility. This leads to high G_t, but at the cost of narrower beams, higher costs of calibration and signal processing for beam-alignment, and lower system yield as more circuitry must be fabricated. Alternatively, the drain efficiency of the PAs, defined as $\eta_D = \frac{\mathrm{P}_t}{\mathrm{P_{DC}}}$ can be improved to increase the ACF. It is to note that this definition of η_D assumes that the active phased array's DC power consumption is mainly due to the power draw from the final output stage, however, there can be further contribution from gain & phase shift control circuitry and pre-amplification/driving stages that limit the actual efficiency enhancement that can be achieved.

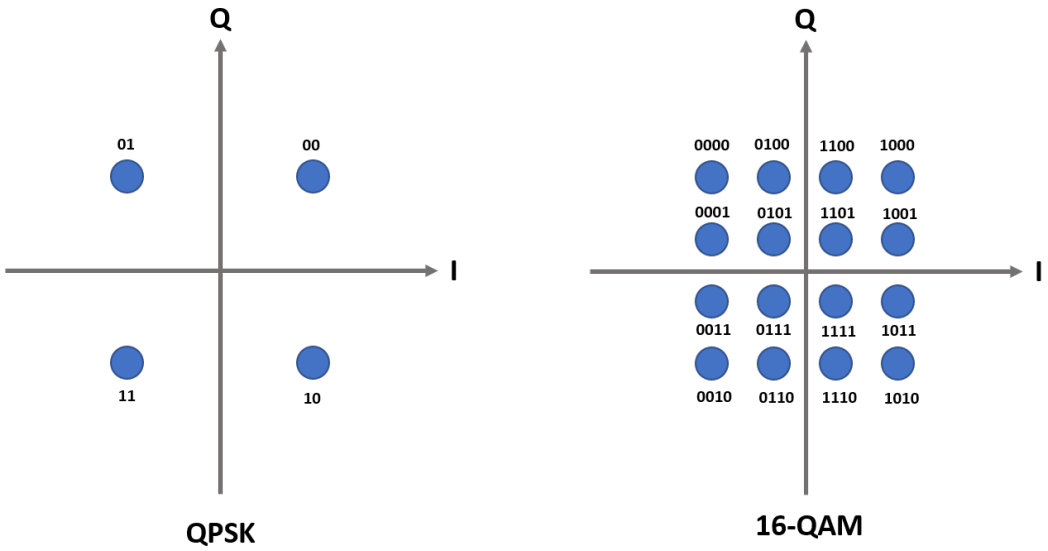

Figure 1.8: Gray-coded constellation of QPSK (left) and 16-QAM (right).

1.6 Addressing Millimeter-Wave Phased Array Challenges with Digital Predistortion

We can define a communication signal, $x(t)$ at passband (RF) as

$$x(t) = \mathbf{Re}\{\tilde{x}(t)e^{j\omega_c t+\beta}\}, \tag{1.9}$$

where $\tilde{x}(t) \in \mathbb{C}$ is the complex baseband envelope of $x(t)$, ω_c is the carrier frequency and β is an arbitrary phase shift. The complex baseband envelope is normally generated from band-limiting the sampled complex baseband envelope, $\tilde{x}[n]$, such that $\tilde{x}(t) = \sum_n \tilde{x}[n]g(t-nT)$, where $g(t)$ is an interpolating filter and T is the sampling period. Modern modulation schemes embed information in the amplitude of $\tilde{x}[n]$, phase of $\tilde{x}[n]$, or a combination of both. These points can take discrete values and map from different combinations of incoming bit-vectors, as shown in Fig. 1.8 for Gray-coded Quadrature Phase-Shift Keying (QPSK) and 16-Quadrature Amplitude Modulation (QAM) schemes.

In single-carrier modulation schemes, the bit-vectors are mapped directly to values of $\tilde{x}[n]$. Advanced modulation schemes such as Orthogonal Frequency-Division Multiplexing (OFDM) have been also developed to allow more flexibility in the waveform and ease the signal processing requirements. In OFDM, L bit-vectors are mapped to L subcarriers,

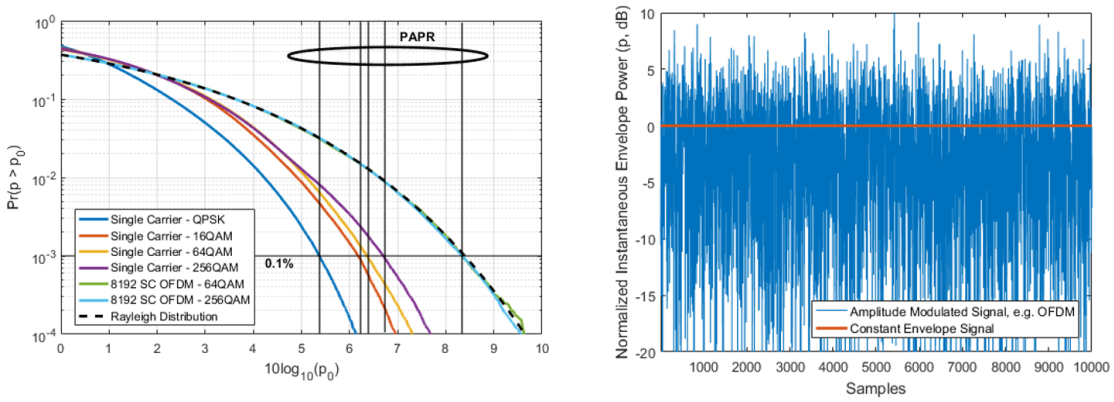

Figure 1.9: CCDF of the instantaneous envelope power, $p[n]$, for different single-carrier and 8192-subcarrier OFDM signals (left). The 0.1% CCDF definition of PAPR is indicated. An example of $p[n]$ (dB) for an OFDM signal over 10000 samples is shown (right).

$\tilde{X}[k]$, and $\tilde{x}[n]$ is generated in an L-length blocks called symbols using an Inverse Fast Fourier Transform (IFFT) as:

$$\tilde{x}[n] = \frac{1}{L} \sum_{k=0}^{L-1} \tilde{X}[k] e^{\frac{j2\pi kn}{L}}. \tag{1.10}$$

$\tilde{x}[n]$ can take on different statistics depending on the modulation scheme. We can define the instantaneous power of the complex envelope, $p[n]$, as:

$$p[n] = \frac{|\tilde{x}[n]|^2}{\mathbf{E}[|\tilde{x}[n]|^2]}. \tag{1.11}$$

Since $x(t)$ is bounded by $\tilde{x}(t)$ and subsequently $\tilde{x}[n]$, this measure is important as it defines the range of power levels above the average power that $x(t)$ can take. The empirical Complementary Cumulative Distribution Function (CCDF) of $p[n]$ for different modulation schemes can be seen in Fig. 1.9 (left).

The CCDF defines the probability that $p[n]$ will be greater than a given value p_0 for a particular signal $\tilde{x}[n]$. Suppose that we want to limit the range of power the $\tilde{x}[n]$ can take without significantly altering the signal. For this, we can find a value of p_0 with low enough probability on the CCDF such that clipping $\tilde{x}[n]$ at all points above this value will not significantly increase the error. This value of p_0 is known as the Peak-to-Average

Power Ratio (PAPR) of the signal, commonly specified as the p_0 which corresponds to the 0.1% CCDF point. As shown in Fig. 1.9 (left), the PAPR increases for higher-order modulation schemes and is the highest for OFDM signals. In fact, high subcarrier-count OFDM will approach in statistics to a Rayleigh distributed random process, which has $\mathbf{Pr}[p > p_0] = e^{-p_0}$ and subsequently a PAPR of around 8.4 dB. To illustrate how much the instantaneous envelope power can vary, a plot of the samples of $p[n]$ for an OFDM signal can be seen in Fig. 1.9 (right).

The PAPR is an important characteristic of a signal since it determines the power level at which a PA can be driven before non-linear distortion becomes significant. For any PA, there is a maximum power at which it can safely output called the saturation level. To output a signal with a given PAPR, the input power must be backed-off by at least the PAPR from the level necessary to reach saturation or else the output will be clipped. Moreover, as the output signal power approaches the saturation level, the PA's input/output relationship will no longer be linear; the output can be amplitude modulated and/or phase modulated depending on the instantaneous input power level. The Amplitude Modulation (AM) induced by the input power is denoted as AM/AM and manifests as gain compression, shown for a typical amplifier in Fig. 1.10 (left). Similarly, the Phase Modulation (PM) induced by the input power is denoted as AM/PM. To avoid the collective effects of AM/AM and AM/PM caused by the non-linear distortion, the input signal must be backed-off even further.

The main issue is that PAs are most efficient near their saturated power level. As the input back-off is increased, the overall efficiency of the amplifier will generally decrease. This is illustrated in Fig. 1.10 (right) for ideal Class A and Class B PAs. It is to note that the efficiency curve reported in Fig. 1.10 represents those of ideal amplifiers. For non-ideal PAs, especially those implemented at mm-wave, the efficiency drop-off is expected to be much higher. One method of efficiency enhancement is to reduce back-off levels at the cost of increased non-linear distortion. If the non-linear distortion can be compensated, then the PA can operate much more efficiently without hampering the signal quality. This is the role of signal predistortion, and more specificially in this work, Digital Predistortion (DPD), in reducing the thermal challenges in mm-wave phased arrays.

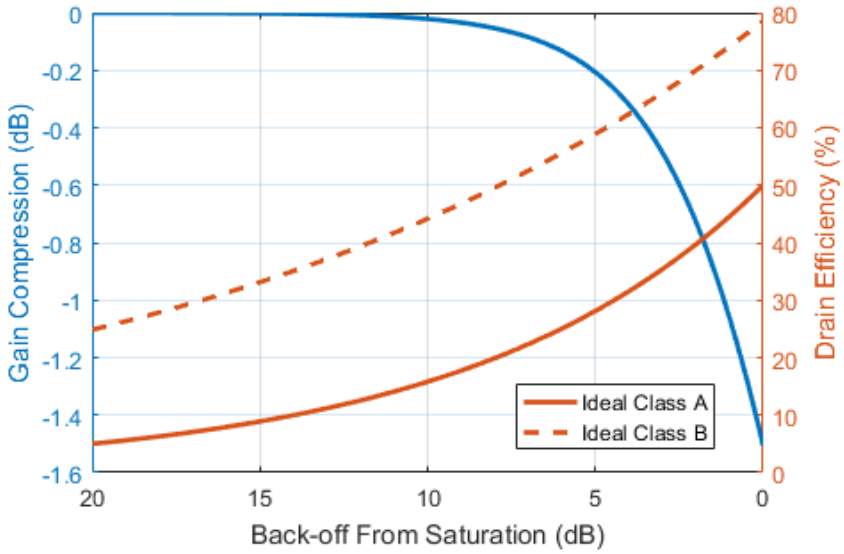

Figure 1.10: Gain compression (dB) vs. input power back-off (dB) from level needed for saturated output (left). Drain efficiency for ideal Class A and Class B amplifiers vs. input power back-off (right).

1.7 Single Power-Amplifier Digital Predistortion Background

There has been much effort undertaken to maximize the power efficiency of Single-Input Single-Output (SISO) 3G and 4G transmitters. Advanced PA power efficiency enhancement techniques, including load and supply modulation based PA systems, have been devised to achieve high power efficiency while driving modulated signals with high PAPR. These include methods such as Doherty PA, Chireix (outphasing) transmitter, and envelope tracking. Yet, the successful practical deployment of these PAs has necessitated major advancement in linearization techniques, particularly digital predistortion, or DPD, to mitigate the significant distortions exhibited by the high efficiency PAs.

As shown in Fig. 1.11, a conventional SISO DPD system consists of three main modules: the Transmission Observation Receiver (TOR), the training algorithm, and the engine. The TOR creates a feedback path from the PA output, $y(t)$, by providing samples of the envelope of $y(t)$, $\tilde{y}[n]$, to the training algorithm. The training algorithm uses $\tilde{y}[n]$ and the original baseband samples, $\tilde{x}[n]$, to determine a function which maps $\tilde{x}[n]$ to its predistorted

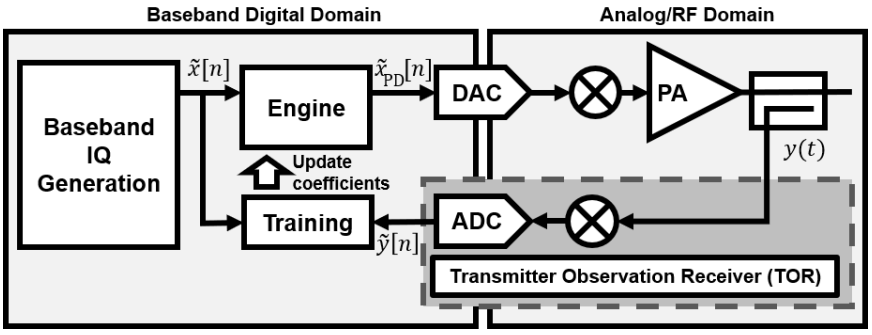

Figure 1.11: Digital predistortion system for a single-PA SISO transmitter.

counterpart $\tilde{x}_{\mathrm{PD}}[n]$ with the goal of making the overall path from $\tilde{x}[n]$ to $y(t)$ linear. This function, known as the DPD function, is then executed in real time using the engine.

RF PAs can be modeled at baseband using non-linear basis-based behavioral functions. These functions are typically derived from Artificial Neural Network (ANN) [19] or Volterra-series based frameworks [20, 21]. One such Volterra-series based model is the Generalized Memory Polynomial (GMP) from [21]. The PA function $f_{\mathrm{PA}}(\cdot)$ is written as a function of an input baseband envelope signal $\tilde{v}[n]$ and delayed taps of $\tilde{v}[n]$ up to memory depth M such that:

$$f_{\mathrm{PA}}(\tilde{v}[n], \tilde{v}[n-1], ..., \tilde{v}[n-M]; \boldsymbol{\alpha}) = \sum_{p=0}^{P-1} \sum_{m=0}^{M_a} \sum_{l=0}^{M_b} \alpha_{p,m,l} |\tilde{v}[n-l-m]|^p \tilde{v}[n-l], \quad (1.12)$$

where $M = M_a + M_b$, $P+1$ is the non-linearity order and $\alpha_{p,m,l}^{(i)}$ is the corresponding basis coefficient in $\boldsymbol{\alpha}$. For simplicity of analysis, since the basis functions in (1.12) all follow the same structure, (1.12) can be written as a summation over the total number of basis functions K for even p as:

$$f_{\mathrm{PA}}(\tilde{v}[n], \tilde{v}[n-1], ..., \tilde{v}[n-M]; \boldsymbol{\alpha}) = \sum_{k=1}^{K} \alpha_k |\tilde{v}[n-m_k]|^{2p_k} \tilde{v}[n-m_k'], \quad (1.13)$$

where m_k, $m_k' \leq M$ are the lags of k'th the polynomial basis and $2p_k+1$ is the nonlinearity order of the k'th basis. This notation can further be simplified by defining

$$f_{\mathrm{PA}}(\tilde{\mathbf{v}}(n); \boldsymbol{\alpha}) = \sum_{k=1}^{K} \alpha_k \psi_k(\tilde{\mathbf{v}}(n)), \quad (1.14)$$

where

$$\tilde{\mathbf{v}}(n) = \left(\tilde{v}[n], \tilde{v}[n-1], \cdots, \tilde{v}[n-M]\right), \tag{1.15}$$

$$\psi_k(\tilde{\mathbf{v}}(n)) = |\tilde{v}[n-m_k]|^{2p_k}\tilde{v}[n-m_k']. \tag{1.16}$$

An example of how a SISO DPD can be trained is as follows. Suppose we have $\tilde{x}[n]$ and the corresponding $\tilde{y}[n]$ phase and time-aligned to $\tilde{x}[n]$. The phase and time alignment can be done using time-domain cross-correlation, frequency-domain phase equalization, or a combination of both (e.g. see [22]). The PA can be represented as a function of the input signal and its delays up to a memory depth M as:

$$\tilde{y}[n] = f_{\text{PA}}(\tilde{\mathbf{x}}(n); \boldsymbol{\alpha}). \tag{1.17}$$

By modeling the nonlinear distortion in $f_{\text{PA}}(\tilde{\mathbf{x}}(n); \boldsymbol{\alpha})$ as additive, 1.17 can be written as

$$\tilde{y}[n] = G\tilde{x}[n] + \tilde{e}(\tilde{\mathbf{x}}(n); \boldsymbol{\beta}), \tag{1.18}$$

where

$$\tilde{e}(\tilde{\mathbf{x}}(n); \boldsymbol{\beta}) = \sum_{k=1}^{K} \beta_k \psi_k(\tilde{\mathbf{v}}(n)), \tag{1.19}$$

and G is the small-signal gain of the system. G can be estimated through separate measurement, or through finding $\hat{G} = \frac{\mathbf{E}[|\tilde{y}[n]|^2}{\mathbf{E}[|\tilde{x}[n]|^2}$. Once $\hat{G}$ is found, the error signal can be approximated by:

$$\tilde{e}(\tilde{\mathbf{x}}(n); \boldsymbol{\beta}) \approx \tilde{y}[n] - \hat{G}\tilde{x}[n]. \tag{1.20}$$

One way that an estimate of $\tilde{e}(\tilde{\mathbf{x}}(n); \boldsymbol{\beta})$ can be found is by finding the estimate for $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, through least-squares fitting. For a length $D + M$ of known $\tilde{x}[n]$ and length D of observed output $\tilde{y}[n]$, this is done by solving the following equation:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{\Psi}^{\text{H}}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\text{H}}\tilde{\mathbf{y}}, \tag{1.21}$$

where

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\psi}(\tilde{\mathbf{x}}(n)) \\ \boldsymbol{\psi}(\tilde{\mathbf{x}}(n-1)) \\ \vdots \\ \boldsymbol{\psi}(\tilde{\mathbf{x}}(n-D+1)) \end{pmatrix}, \tag{1.22}$$

$$\boldsymbol{\psi}(\tilde{\mathbf{x}}(n)) = \left(\psi_1(\tilde{\mathbf{x}}(n)), \psi_2(\tilde{\mathbf{x}}(n)), \cdots, \psi_K(\tilde{\mathbf{x}}(n))\right), \tag{1.23}$$

$$\tilde{\mathbf{e}} = \begin{pmatrix} \tilde{y}[n] - \hat{G}\tilde{x}[n] \\ \tilde{y}[n-1] - \hat{G}\tilde{x}[n-1] \\ \vdots \\ \tilde{y}[n-D+1] - \hat{G}\tilde{x}[n-D+1] \end{pmatrix}, \tag{1.24}$$

and $\mathbf{\Psi}^{\mathrm{H}}$ denotes the Hermitian transpose of matrix $\mathbf{\Psi}$. The estimated distortion is then given by $\tilde{e}(\tilde{\mathbf{x}}(n); \hat{\boldsymbol{\beta}})$ and the estimated PA output $\hat{\tilde{y}}[n]$ can then be written as

$$\hat{\tilde{y}}[n] = \hat{G}\tilde{x}[n] + \tilde{e}(\tilde{\mathbf{x}}(n); \hat{\boldsymbol{\beta}}). \tag{1.25}$$

If f_{PA} is weakly nonlinear, then the following approximation can be made if a small perturbation signal $\epsilon[n]$ is added to $\tilde{x}[n]$:

$$\tilde{x}_{\mathrm{PD}}[n] = \tilde{x}[n] + \epsilon[n] \tag{1.26}$$

$$f_{\mathrm{PA}}(\tilde{\mathbf{x}}_{\mathrm{PD}}(n); \boldsymbol{\alpha}) = G(\tilde{x}[n] + \epsilon[n]) + \tilde{e}(\tilde{\mathbf{x}}_{\mathrm{PD}}(n); \boldsymbol{\beta}) \tag{1.27}$$

$$\approx G\tilde{x}[n] + G\epsilon[n] + \tilde{e}(\tilde{\mathbf{x}}(n); \boldsymbol{\beta}). \tag{1.28}$$

Hence, for small magnitudes of non-linear distortion, we can generate the predistortion signal $\tilde{x}_{\mathrm{PD}}[n]$ to cancel $\tilde{e}(\tilde{\mathbf{x}}(n); \boldsymbol{\beta})$ directly as

$$\tilde{x}_{\mathrm{PD}}[n] = \tilde{x}[n] - \frac{\tilde{e}(\tilde{\mathbf{x}}(n); \hat{\boldsymbol{\beta}})}{\hat{G}}. \tag{1.29}$$

The generated predistortion signal is likely to expand (i.e. the PAPR of $\tilde{x}_{\mathrm{PD}}[n]$ is expected to increase) to compensate for gain compression. This change in the statistics of the input signal will subsequently result in additional non-linear distortion at the PA output which would not have been present during the initial estimation period. It is therefore expected that multiple iterations of the predistortion signal upload, TOR output signal capture, and residual error model estimation be done before converging at the final DPD function.

The AM/AM and AM/PM plots in Fig. 1.12 illustrate typical behavior of the various signals in the DPD system after convergence. As can be seen from Fig. 1.12 (left), the predistortion signal $\tilde{x}_{\mathrm{PD}}[n]$, expands the amplitude of original input signal $\tilde{x}[n]$. When it is cascaded with the compression of the PA, the overall transfer function from $\tilde{x}[n]$ to $\tilde{y}[n]$ becomes linear. This is also observed in the phase response as shown in AM/PM plot in Fig. 1.12 (right). After DPD, the PA should be operating with lower input power backoff (and thus higher efficiency) to achieve the same, if not better, signal quality as if it were operated at a higher backoff but without DPD applied.

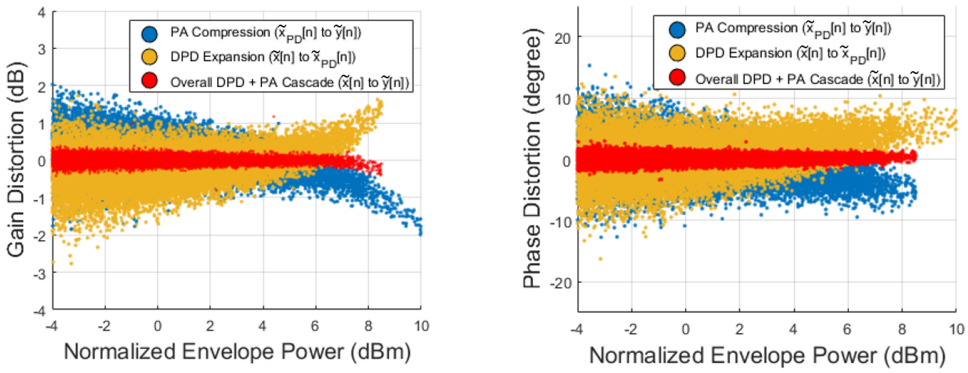

Figure 1.12: Amplitude modulation induced by the input amplitude (AM/AM, left) and Phase modulation induced by input amplitude (AM/PM, right) for various points in the DPD system.

One must note that there are limitations in the amount of compression that the DPD can correct. Higher amount of compression would require the predistortion signal to correspondingly expand further while maintaining the same input average power. At some point, the expansion will be limited by the saturated output power of the PA. One should also note that DPD will be most effective for well-designed PAs with low amount of initial AM/AM and AM/PM distortion. Correcting for significant AM/AM and AM/PM typically requires more basis coefficients, which can lead to difficulty in training and efficiently generating the predistortion signal.

Chapter 2

Digital Predistortion for Single-User RF Beamforming

To simultaneously meet 5G mobility requirements and overcome the high path loss inherent in mm-wave, the transition from conventional single-antenna transceivers to ones based on large-scale antenna arrays has become a rapidly growing topic of interest. Several arrangements of large-scale array-based transmitters have been devised where the beamforming feature is either realized in analog/RF, digital or a hybrid of both. RF beamforming based architectures using deep sub-micron semiconductor technologies has so far been the most adopted [23, 24]. These RF beamforming arrays typically consist of analog gain and phase shifters, and array of amplifiers with each driving a single element. A diagram of a typical RF beamforming array and example of a first-generation commercial 64-element implementation are shown in Fig. 2.1.

2.1 Towards Digital Predistortion of Millimeter-wave RF Beamforming Arrays

2.1.1 Prior works

There have been some attempts to extend the application of DPD to multiple-antenna transmitter arrays. These can be classified in 2 distinct categories: those based on SISO-based modeling approaches (inputs are forward waves) and those that use dual-input modeling approaches (inputs are forward and reverse waves). The first of the categories is the

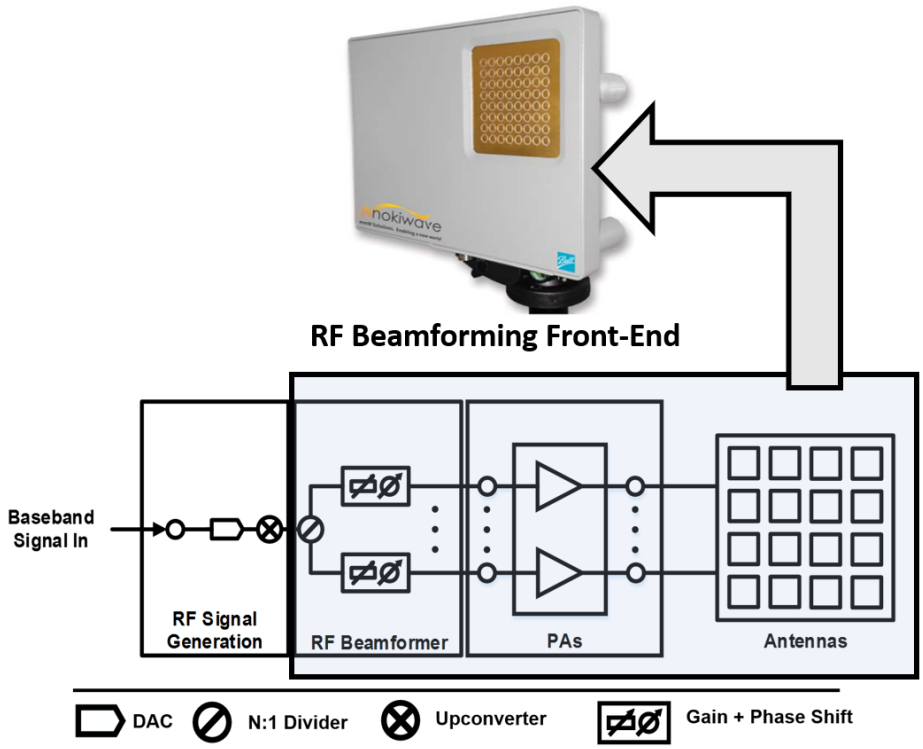

Figure 2.1: Block diagram of a typical RF beamforming front-end and a 64-Element RF Beamforming Phased Array at 28 GHz [5].

SISO approach, which covers works in [25–29]. In [25], all PA outputs in an RF beamforming array are observed. A single polynomial-based DPD is derived for the array based on the collective outputs and compared to the case where only a single PA is modeled. In [26, 28], a DPD is derived based on minimizing the nonlinear distortion at a particular angle in the far-field and [27] experimentally verifies the method by linearizing a 2x2 array operating at sub-6 GHz driven by 10 MHz modulated signals. Lastly, [29] applies a SISO model to linearize a hybrid beamforming array in simulation.

In contrast, a dual-input model is used for DPD in [30, 31]. The dual-input model is extracted from the individual PAs first and then used in conjunction with a cross-talk and mismatch model to linearize a 4-element digital beamforming array operating at 2.1 GHz

driven by 5 MHz modulated signals. This dual-input model has also been used in other works [32–35] in a similar manner to predict the behavior of RF beamforming arrays that are implemented with the modeled PAs.

While the reported works have been successful in demonstrating the applicability of DPD in arrays, challenges specific to mm-wave DPD have yet to be addressed. The existing SISO works do not incorporate non-ideal effects such as mutual coupling, antenna impedance load modulation and phase-shifter gain variation that would also be expected in mm-wave arrays. The dual-input works require measurement of the individual PAs, which would be difficult for mm-wave arrays given the high number of PA-elements and the compact size of the array. Most mm-wave RF beamforming arrays do not have access to the individual PA outputs, and hence it would be difficult to obtain the feedback signals needed for training the dual-input models. In both the SISO and dual-input cases, the reported measurement results are limited to implementations centered at sub-6 GHz, small array sizes, and relatively narrow-band signals.

2.1.2 Novelty of this work

In this work, an RF beamforming array system – which includes the phase shifters, PAs, and channel/antennas – is modeled altogether as an equivalent SISO model. Non-ideal effects such as uncalibrated phase shifters and PA load impedance modulation are taken into account using sets of SISO function coefficients. The SISO model is then used to linearize a commercial 64-element and 4-element RF beamforming array centered at 28 GHz and driven by 800 MHz modulated signals. The measurement results indicate that the steering range in which these equivalent SISO model coefficients sets are valid over can be extended by multiplexing between different coefficient sets and use of appropriate receiver equalization.

2.2 SISO Model for RF Beamforming Arrays

An ideal RF beamforming system with non-linear PAs is shown in Fig. 2.2 and mathematically described as follows. Let $\tilde{x}[n]$ be the sampled baseband-equivalent envelope to the RF beamformer. The beamformer is modeled by the beamforming weight vector $\mathbf{w} \in \mathbb{C}^{N\times 1}$ and the input signal to the i'th PA is $\tilde{u}_i[n]$, where $i \in \{1, 2, ..., N\}$ is the number of elements in the array and:

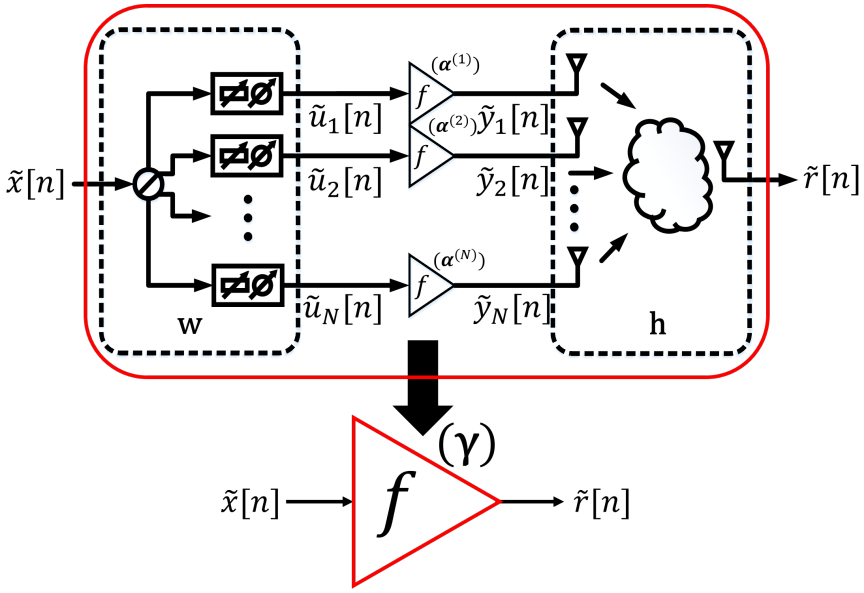

Figure 2.2: Behavioral modeling of RF beamforming PA array using an equivalent SISO describing function.

$$\tilde{\mathbf{u}}[n] = \begin{pmatrix} \tilde{u}_1[n] \\ \tilde{u}_2[n] \\ \vdots \\ \tilde{u}_i[n] \\ \vdots \\ \tilde{u}_N[n] \end{pmatrix} = \mathbf{w}\tilde{x}[n] = \begin{pmatrix} w_1\tilde{x}[n] \\ w_2\tilde{x}[n] \\ \vdots \\ w_i\tilde{x}[n] \\ \vdots \\ w_N\tilde{x}[n] \end{pmatrix}. \tag{2.1}$$

Each of the N PAs are modeled using a baseband-equivalent nonlinear basis-based behavioral function, $f(\tilde{\mathbf{v}}[n]; \boldsymbol{\alpha}^{(i)})$ up to memory depth M, where $\boldsymbol{\alpha}^{(i)} \in \mathbb{C}^{K\times 1}$ is a vector of the K coefficients for the i'th PA. Assuming the PAs are modelled based on GMP and written in the same form as described in (1.12) and (1.13) from Section 1.7, then:

$$f(\tilde{\mathbf{v}}[n]; \boldsymbol{\alpha}^{(i)}) = \sum_{k=1}^{K} \alpha_k^{(i)} |\tilde{v}[n - m_k]|^{2p_k} \tilde{v}[n - m'_k]. \tag{2.2}$$

The output of the i'th PA in the array can then be written as:

$$\tilde{y}_i[n] = \sum_{k=1}^{K} \alpha_k^{(i)} |\tilde{u}_i[n-m_k]|^{2p_k} \tilde{u}_i[n-m_k'] \tag{2.3}$$

$$= \sum_{k=1}^{K} \alpha_k^{(i)} |w_i|^{2p_k} |\tilde{x}[n-m_k]|^{2p_k} w_i \tilde{x}[n-m_k']. \tag{2.4}$$

The user received far-field signal, $\tilde{r}[n]$, across a channel modeled as $\mathbf{h} \in \mathbb{C}^{1\times N}$ is found as:

$$\tilde{r}[n] = \sum_{i=1}^{N} h_i \tilde{y}_i[n] \tag{2.5}$$

$$= \sum_{i=1}^{N} h_i w_i \sum_{k=1}^{K} \alpha_k^{(i)} |w_i|^{2p_k} |\tilde{x}[n-m_k]|^{2p_k} \tilde{x}[n-m_k'] \tag{2.6}$$

$$= \sum_{k=1}^{K} \gamma_k |\tilde{x}[n-m_k]|^{2p_k} \tilde{x}[n-m_k'], \tag{2.7}$$

where $\gamma_k = \sum_{i=1}^{N} \alpha_k^{(i)} h_i w_i |w_i|^{2p_k}$. From ((2.7)) it is evident that the relationship between $\tilde{x}[n]$ and $\tilde{r}[n]$ has the same form of the individual PA behavioral function ((2.4)) but with different model coefficients, as illustrated in Fig. 2.2. This transfer function is denoted by $f(\tilde{\mathbf{v}}[n];\boldsymbol{\gamma})$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_K)$ and known as the SISO model for RF beamforming PA arrays.

2.3 SISO Model for Ideal Millimeter-wave RF Beamforming

For mm-wave arrays, the channel is often characterized as a single Line-of-Sight (LOS) path. In this case, $\mathbf{h}$ is determined by phase shifts depending on the angular position of the user relative to the broadside of the array. This position is specified on the elevation and azimuthal axis as θ and ϕ respectively and is related to the LOS channel vector, $\mathbf{h}^{\text{LOS}}$ by:

$$h_i^{\text{LOS}}(\theta,\phi) = e^{-j\frac{2\pi}{\lambda}\sin(\theta)(d_i^x\cos(\phi)+d_i^y\sin(\phi))}, \tag{2.8}$$

where (d_i^x,d_i^y) is the position of the i'th radiator along the x- and y-axis respectively of the array plane relative to a reference antenna.

A common precoding method is Maximum Ratio Transmission (MRT), which selects $\mathbf{w}$ such that the SNR is maximized [36]. To realize MRT in the LOS case, $\mathbf{w}^{\mathrm{MRT}}(\theta,\phi) = (\mathbf{h}^{\mathrm{LOS}}(\theta,\phi))^{\mathrm{H}}$ is selected, where $\bar{\mathbf{x}}^{\mathrm{H}}$ denotes the Hermitian-transpose of a vector $\bar{\mathbf{x}}$. In AF terms, MRT precoding is equivalent to steering the beam towards the user at (θ,ϕ). Noting that $|w_i^{\mathrm{MRT}}|^{2p} = 1$, the received signal at observation angle (θ_o, ϕ_o) with beam-steering direction (θ, ϕ), $\tilde{r}(\theta_o, \phi_o, \theta, \phi)[n]$, is found by setting $h_i = h_i^{\mathrm{LOS}}(\theta_o, \phi_o)$ and $w_i = w_i^{\mathrm{MRT}}(\theta, \phi)$ and is written as:

$$\tilde{r}(\theta_o, \phi_o, \theta, \phi)[n] = \sum_{k=1}^{K} \gamma_k(\theta_o, \phi_o, \theta, \phi)|\tilde{x}[n - m_k]|^{2p_k}\tilde{x}[n - m_k'], \tag{2.9}$$

where

$$\gamma_k(\theta_o, \phi_o, \theta, \phi) = \sum_{i=1}^{N} \alpha_k^{(i)} h_i^{\mathrm{LOS}}(\theta_o, \phi_o) w_i^{\mathrm{MRT}}(\theta, \phi). \tag{2.10}$$

Of note are a few interesting insights of the SISO model in a mm-wave system. If all the individual PA behavioral function coefficients are identical (i.e. $\alpha_k^{(i)} = \alpha_k$), then every basis of the SISO model experiences multiplication by the same scalar factor $\beta(\theta_o, \phi_o, \theta, \phi)$, i.e.:

$$\gamma_k(\theta_o, \phi_o, \theta, \phi) = \alpha_k \beta(\theta_o, \phi_o, \theta, \phi), \tag{2.11}$$

where

$$\beta(\theta_o, \phi_o, \theta, \phi) = \sum_{i=1}^{N} h_i^{\mathrm{LOS}}(\theta_o, \phi_o) w_i^{\mathrm{MRT}}(\theta, \phi). \tag{2.12}$$

Since the scaling factor is identical for linear and non-linear components, it is expected that uncorrected Out-of-band (OOB) emissions for RF beamforming arrays with similar PAs will approximately follow the same radiation pattern as the in-band spectrum.

In the user direction, $(\theta_o, \phi_o) = (\theta, \phi)$ and thus the user-received signal is given by ((2.9)), where

$$\gamma_k(\theta_o, \phi_o, \theta, \phi) = \gamma_k(\theta, \phi, \theta, \phi) = \sum_{i=1}^{N} \alpha_k^{(i)}. \tag{2.13}$$

From ((2.13)) it can be seen that the γ_k and the user-received signal are no longer a function of the angular position of the user. Subsequently, in the ideal case, a single SISO nonlinear distortion model, $f(\tilde{\mathbf{v}}[n]; \boldsymbol{\gamma})$ can be used to model the PA array across any choice of θ and ϕ. The SISO array model coefficients are, interestingly, the sum of the individual PA model coefficients.

2.4 SISO Model for Non-Ideal Millimeter-wave RF Beamforming

In the non-ideal scenario, the overall PA array model in the user's direction can change depending on the particular choice of θ and ϕ by the following mechanisms. At the output of the PA, mutual coupling between antenna elements can cause each PA to observe a different active reflection coefficient. Additionally, uncorrected phase shifter gain variation can cause the PAs to be driven at different input power. The PA load modulation and phase shifter gain variation effects are first examined in more detail before formulating the equivalent non-ideal SISO array model.

2.4.1 Power Amplifier Load Impedance Variation vs. Steering Angle

In antenna arrays, the coupling between elements can cause the impedance seen at each antenna port to change depending on the phasing of the port excitations. It is determined by the active reflection coefficient, and is defined for the i'th element in an N-antenna array shown in Fig. 2.3 using the scattering parameters of the N-port network as:

$$\Gamma_i = \frac{a_2^{(i)}}{b_2^{(i)}} = \frac{\sum_{j=1}^{N} S_{i,j} b_2^{(j)}}{b_2^{(i)}}, i \in \{1, 2, ..., N\}. \tag{2.14}$$

As an example to illustrate the effect of this impedance variation, the linear patch array shown in Fig. 2.4 was simulated and the S-Parameters versus frequency are reported in Fig. 2.5. The array consists of 4 linearly polarized patches built on top of Rogers 4350B substrate. From the simulated S-parameters of the patch antenna, it can be seen that the adjacent element coupling is around -20 dB and diagonal-element coupling is less than -30 dB. The array's 4-ports are then stimulated with the phasing required to beamsteer to broadside ($\theta = 0°$), $\theta = 30°$ and $\theta = 60°$ at 28 GHz. The active reflection coefficients for the 4 arrays are plotted in Fig. 2.6. From Fig. 2.6, it is evident that both the amplitude and phase of the active reflection coefficient change with steering angle. This change in the reflection coefficient versus steering angle can affect the non-linear behavior of the PA.

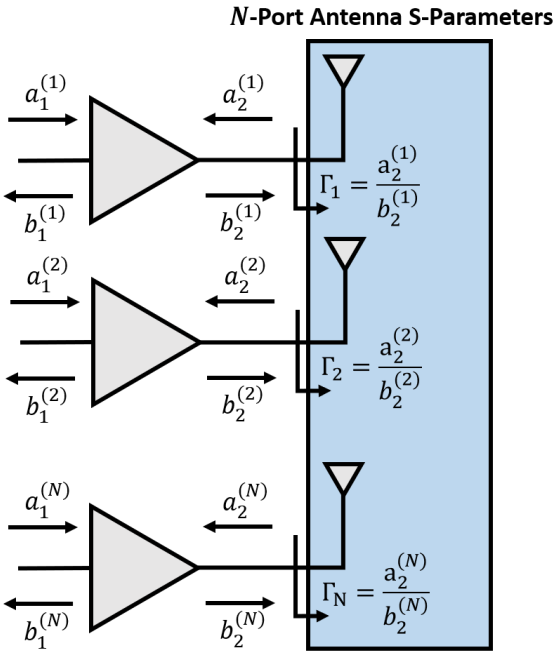

Figure 2.3: Active RF beamforming array with N PAs driving N antennas.

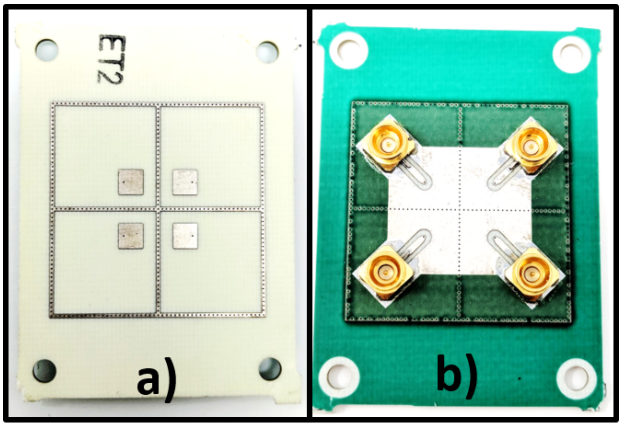

Figure 2.4: The 4-element linear patch antenna. a) Top view. b) Bottom view.

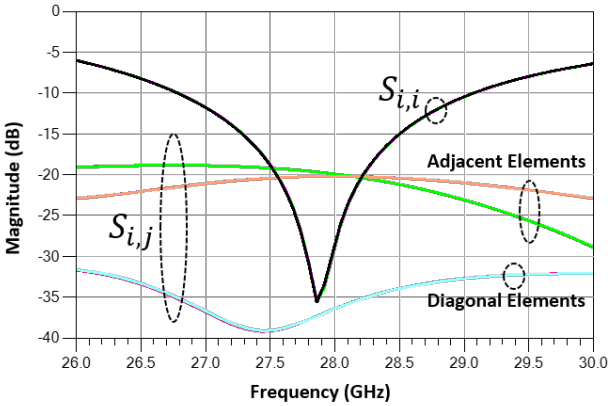

Figure 2.5: Simulated antenna S-parameters showing the magnitude of the diagonals ($S_{i,i}$) and magnitude of cross components ($S_{i,j}$).

2.4.2 Phase Shifter Amplitude Variation vs. Steering Angle

Another non-ideal effect that can cause the non-linear behavior of the array to depend on the beam-steering angle is the gain variation of the phase shifter across phase shift settings. In digital phase shifters, each phase setting corresponds to a specific electrical length that is achieved by enabling or disabling different signal paths. Due to the possibility for differences in losses in each path, the independence between the phase-setting and gain of the phase shifter is not necessarily ensured. This problem is illustrated with measurement results of a commercial 5-bit digital phase shifter at 28 GHz shown in Fig. 2.7 (left). The 5-bit phase shifter has a bit for each of the 11.25°, 22.5°, 45°, 90°, and 180° paths, and each bit is 0 if the phase path is not activated and 1 if it is activated. The concatenated phase shifter bits correspond to a 5-bit code, and the measured gain normalized to the 0 code is plotted. It can be seen from Fig. 2.7 that depending on the phase shifter code, the gain can vary up to 0.8 dB.

Next, an array of 4 phase shifters is considered. Each phase shifter is given the gain vs. phase response found in Fig. 2.7 (left), which were found after averaging the measured results of 4 samples of a 5-bit commercial beamforming IC. The phase shifter values are then set using phase codes that correspond to steering a 2x2 $\frac{\lambda}{2}$-spaced array to angles between $-60° \leq \theta \leq 60°$. The power at the output of the non-ideal phase shifters are summed and plotted across the steering angles and shown in Fig. 2.7 (right). From the plot, a variation of about 0.375 dB can be seen in the total power at the output of the phase-shifters. Since the input power to the PAs determine the amount of compression

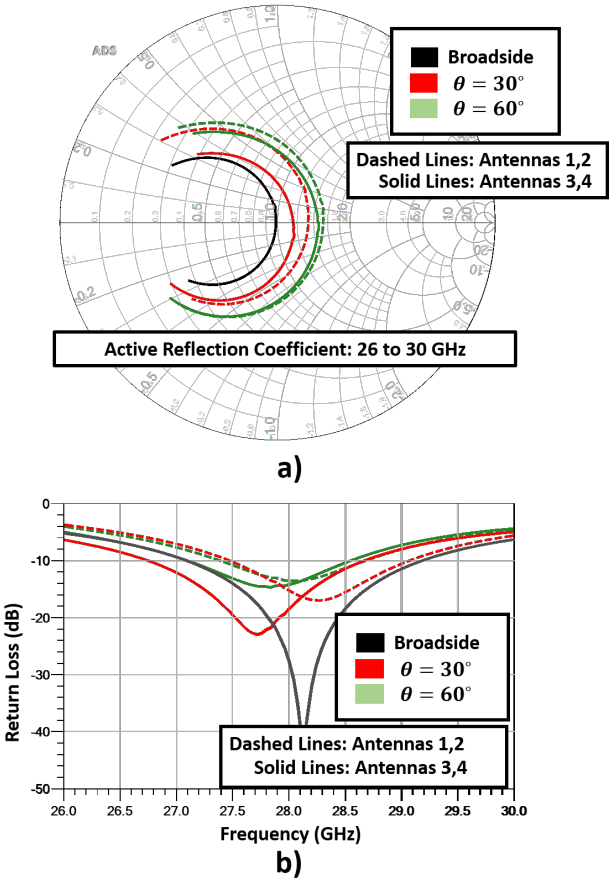

Figure 2.6: Active reflection coefficient for the linear patch array for the array steered at broadside, $\theta = 30°$, and $\theta = 60°$. The active reflection coefficient is plotted with magnitude and phase in a) and on a dB scale in b).

experienced by the array, the nonlinear behavior of the array is expected to change due to the phase shifter gain variation. Note that this issue is implementation dependent, as there are many ways of realizing an RF phase shifter. The phase-dependent gain is not expected to be as significant in mixer or Local Oscillator (LO) based phase shifter implementations.

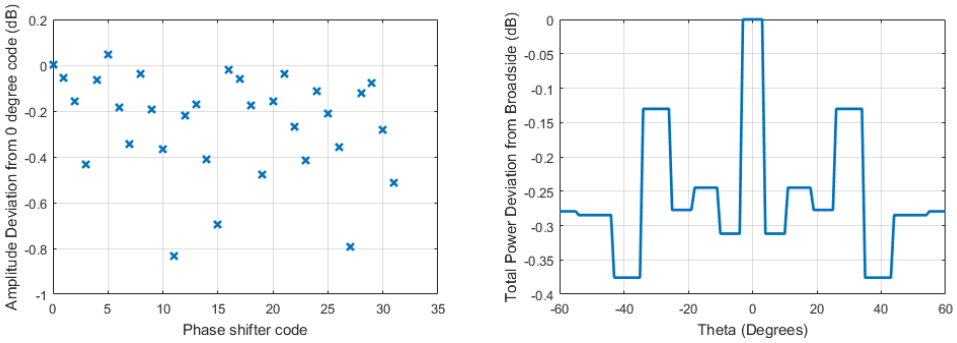

Figure 2.7: Measured phase shifter amplitude dependency on phase shifter setting (left). Predicted total output power variation with steering angle in a 4-element array (right).

2.4.3 Non-ideal SISO DPD Formulation

In the literature, dual-input PA models have been reported to account for the effects of PA load impedance variation in individual PAs [37] and in arrays [30, 32–35]. These works include both $a_1[n]$ and $a_2[n]$ in the non-linear basis functions. For example, the formulation in [35] has the output of a PA modeled with incident waves $a_1[n]$ and $a_2[n]$ as:

$$g_{\text{PA}}(\mathbf{a}_1(n), \mathbf{a}_2(n)) = g_1(\mathbf{a}_1(n)) + g_2(\mathbf{a}_1(n), \mathbf{a}_2(n)) + g_3(\mathbf{a}_1(n), \mathbf{a}_2^*(n)), \tag{2.15}$$

where

$$\mathbf{a}_1(n) = \left(a_1[n], a_1[n-1], \cdots, a_1[n-M]\right), \tag{2.16}$$

$$\mathbf{a}_2(n) = \left(a_2[n], a_2[n-1], \cdots, a_2[n-M]\right), \tag{2.17}$$

and

$$g_1(\mathbf{a}_1(n)) = \sum_{m=0}^{M} \sum_{p=0}^{P} \alpha_{p,m,1} a_1[n-m]|a_1[n-m]|^p, \tag{2.18}$$

$$g_2(\mathbf{a}_2(n), \mathbf{a}_1(n)) = \sum_{m=0}^{M} \sum_{p=0}^{P} \alpha_{p,m,1} a_2[n-m]|a_1[n-m]|^p, \tag{2.19}$$

$$g_3(\mathbf{a}_2^*(n), \mathbf{a}_1(n)) = \sum_{m=0}^{M} \sum_{p=0}^{P} \alpha_{p,m,1} a_2^*[n-m] \frac{(a_1[n-m])^2}{|a_1[n-m]|^2} |a_1[n-m]|^p. \tag{2.20}$$

In the following analysis, we will show that (2.15) can be reduced to a SISO model under the following assumptions:

1. RF beamforming is used, which makes the i'th PA input, $a_1^{(i)}[n]$, a scaled version of a common input $a_1[n]$, i.e.
$$a_1^{(i)}[n] = a_1[n]w_i, \tag{2.21}$$
where $w_i \in \mathbb{C}$ is the RF beamforming coefficient for the i'th PA.

2. Each PA output can be modeled as a weakly non-linear function of its input $a_1^{(i)}[n]$, i.e.
$$b_2^{(i)}[n] = (\eta^{(i)} * a_1^{(i)})[n] + \epsilon^{(i)}[n], \tag{2.22}$$
where $\boldsymbol{\eta}^{(i)} \in \mathbb{C}^M$ is a filter corresponding to the linear distortion and the $\epsilon^{(i)}[n]$ are non-linear error terms which are considered small compared to the linear terms.

With this SISO formulation, the modeling complexity of RF beamforming arrays with antenna load modulation is subsequently reduced.

First, we define the basis in (2.15) to include cross-memory terms and only odd non-linearity orders. We assume, without loss of generality, that the same set of dual-input non-linear basis functions are used for each of the PAs in the array. The corresponding basis can have different dual-inputs for each of the PAs and is decomposed into 3 types of basis functions, $\psi_{1,k_1}(\mathbf{a}_1^{(i)}(n))$, $\psi_{2,k_2}(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n))$, and $\psi_{3,k_3}(\mathbf{a}_2^{*(i)}(n), \mathbf{a}_1^{(i)}(n))$, defined as:

$$\psi_{1,k_1}(\mathbf{a}_1^{(i)}(n)) = a_1^{(i)}[n - m'_{1,k_1}]|a_1^{(i)}[n - m_{1,k_1}]|^{2p_{1,k_1}}, \tag{2.23}$$
$$\psi_{2,k_2}(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n)) = a_2^{(i)}[n - m'_{2,k_2}]|a_1^{(i)}[n - m_{2,k_2}]|^{2p_{2,k_2}}, \tag{2.24}$$
$$\psi_{3,k_3}(\mathbf{a}_2^{*(i)}(n), \mathbf{a}_1^{(i)}(n)) = a_2^{*(i)}[n - m'_{3,k_3}]\frac{(a_1^{(i)}[n - m'_{3,k_3}])^2}{|a_1^{(i)}[n - m'_{3,k_3}]|^2}|a_1^{(i)}[n - m_{3,k_3}]|^{2p_{3,k_3}}, \tag{2.25}$$

where k_u is the basis index, $2p_{u,k_u} + 1$ is the nonlinearity order and m_{u,k_u}, m'_{u,k_u} are the lags of the k_u'th basis in $\psi_{u,k_u}(\cdot)$ for $u = 1, 2, 3$. The i'th PA can be individually described using vectors of model coefficients, $\boldsymbol{\alpha}_1^{(i)} \in \mathbb{C}^{K_1}$, $\boldsymbol{\alpha}_2^{(i)} \in \mathbb{C}^{K_2}$, $\boldsymbol{\alpha}_3^{(i)} \in \mathbb{C}^{K_3}$ and hence the overall model for the output of the i'th PA can be written as:

$$f_{\text{PA}}(\mathbf{a}_1^{(i)}(n), \mathbf{a}_2^{(i)}(n); \boldsymbol{\alpha}_1^{(i)}, \boldsymbol{\alpha}_2^{(i)}, \boldsymbol{\alpha}_3^{(i)}) = f_1(\mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_1^{(i)}) \tag{2.26}$$
$$+ f_2(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_2^{(i)}) \tag{2.27}$$
$$+ f_3(\mathbf{a}_2^{*(i)}(n), \mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_3^{(i)}), \tag{2.28}$$

where

$$f_1(\mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_1^{(i)}) = \sum_{k_1=1}^{K_1} \alpha_{1,k_1}^{(i)} \psi_{1,k_1}(\mathbf{a}_1^{(i)}(n)), \tag{2.29}$$

$$f_2(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_2^{(i)}) = \sum_{k_2=1}^{K_2} \alpha_{2,k_2}^{(i)} \psi_{2,k_2}(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n)), \tag{2.30}$$

$$f_3(\mathbf{a}_2^{*(i)}(n), \mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_3^{(i)}) = \sum_{k_3=1}^{K_3} \alpha_{3,k_3}^{(i)} \psi_{3,k_3}(\mathbf{a}_2^{*(i)}(n), \mathbf{a}_1^{(i)}(n)). \tag{2.31}$$

Next we consider each of the $f_1(\cdot)$, $f_2(\cdot)$, and $f_3(\cdot)$ terms separately. From the RF beamforming assumption made in (2.21), it follows that $a_1^{(i)}[n] = a_1[n]w_i$ and thus (2.29) can be written as:

$$\begin{aligned} f_1(\mathbf{a}_1^{(i)}(n); \boldsymbol{\alpha}_1^{(i)}) &= f_1(\mathbf{a}_1(n)w_i; \boldsymbol{\alpha}_1^{(i)}) \\ &= \sum_{k_1=1}^{K_1} \alpha_{1,k_1}^{(i)} \psi_{1,k_1}(\mathbf{a}_1(n)w_i) \\ &= \sum_{k_1=1}^{K_1} \zeta_{1,k_1}^{(i)} a_1[n - m'_{1,k_1}] |a[n - m_{1,k_1}]|^{2p_{1,k_1}}, \end{aligned} \tag{2.32}$$

where

$$\zeta_{1,k_1}^{(i)} = \alpha_{1,k_1}^{(i)} |w_i|^{2p_{1,k_1}} w_i. \tag{2.33}$$

Thus the first term $f_1(\mathbf{a}_1^{(i)}[n]; \boldsymbol{\alpha}_1^{(i)})$ can be written in terms of $a_1[n]$ only.

For $f_2(\cdot)$, we note that the relationship between $a_2^{(i)}[n]$ and $b_2^{(i)}[n]$ can be represented in the time-domain using L-length FIR filters $\lambda_{i,j}[n]$ as is done in [30], written as:

$$a_2^{(i)}[n] = \sum_{j=1}^{N} (\lambda_{i,j} * b_2^{(j)})[n]. \tag{2.34}$$

With the weakly-nonlinear assumption in (2.22),

$$b_2^{(j)}[n] = (\eta^{(j)} * a_1^{(j)})[n] + \epsilon^{(j)}[n], \tag{2.35}$$

and subsequently, (2.34) can be written as:

$$a_2^{(i)}[n] = \sum_{j=1}^{N} ((\beta_{i,j} * a_1^{(j)})[n] + \Delta_{i,j}[n]), \tag{2.36}$$

where

$$\beta_{i,j}[n] = (\lambda_{i,j} * \eta^{(j)})[n], \tag{2.37}$$

$$\Delta_{i,j}[n] = (\lambda_{i,j} * \epsilon^{(j)})[n]. \tag{2.38}$$

Substituting (2.36) into (2.24) we get:

$$\begin{aligned}
&\psi_{2,k_2}(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n)) \\
&\quad = \psi_{2,k_2}(\sum_{j=1}^{N} \left((\boldsymbol{\beta}_{i,j} * \mathbf{a}_1^{(j)})(n) + \boldsymbol{\Delta}_{i,j}(n) \right), \mathbf{a}_1^{(i)}(n)) \\
&\quad = \psi_{2,k_2}(\sum_{j=1}^{N} (\boldsymbol{\beta}_{i,j} * \mathbf{a}_1^{(j)})(n), \mathbf{a}_1^{(i)}(n)) + \psi_{2,k_2}(\sum_{j=1}^{N} \boldsymbol{\Delta}_{i,j}(n), \mathbf{a}_1^{(i)}(n)) \\
&\quad \approx \psi_{2,k_2}(\sum_{j=1}^{N} (\boldsymbol{\beta}_{i,j} * \mathbf{a}_1^{(j)})(n), \mathbf{a}_1^{(i)}(n))
\end{aligned} \tag{2.39}$$

with the approximation made as the terms with $\boldsymbol{\Delta}_{i,j}(n)$ in the expansion are second-order effects consisting of non-linear error terms, which are already assumed small relative to the linear terms in (2.22). Applying the RF beamforming assumption, (2.39) becomes:

$$\begin{aligned}
\psi_{2,k_2}(\mathbf{a}_2^{(i)}(n), \mathbf{a}_1^{(i)}(n)) &= \psi_{2,k_2}(\sum_{j=1}^{N} (\boldsymbol{\beta}_{i,j} * \mathbf{a}_1)(n) w_j, \mathbf{a}_1(n) w_i) \\
&= \psi_{2,k_2}((\boldsymbol{\beta}_i * \mathbf{a}_1)(n), \mathbf{a}_1(n) w_i),
\end{aligned} \tag{2.40}$$

where

$$\beta_i[n] \triangleq \sum_{j=1}^{N} \beta_{i,j} w_j. \tag{2.41}$$

After expanding (2.40) and substituting back into (2.30), we get:

$$\begin{aligned}
f_2((\boldsymbol{\beta}_i * \mathbf{a}_1)(n), \mathbf{a}_1(n) w_i; \boldsymbol{\alpha}_2^{(i)}) &= \sum_{k_2=1}^{K_2} \alpha_{2,k_2}^{(i)} \psi_{2,k_2}((\boldsymbol{\beta}_i * \mathbf{a}_1)(n), \mathbf{a}_1(n) w_i) \\
&= \sum_{k_2=1}^{K_2} (\zeta_{2,k_2}^{(i)} * a_1)[n - m'_{2,k_2}] |a_1[n - m_{2,k_2}]|^{2p_{2,k_2}},
\end{aligned} \tag{2.42}$$

where

$$\zeta_{2,k_2}^{(i)}[n] = \alpha_{2,k_2}^{(i)} |w_i|^{2p_{2,k_2}} \beta_i[n]. \tag{2.43}$$

Similarly, after applying the same steps for $f_3(\cdot)$:

$$f_3((\boldsymbol{\beta}_i * \mathbf{a}_1)^*(n), \mathbf{a}_1(n)w_i; \boldsymbol{\alpha}_2^{(i)}) = \sum_{k_3=1}^{K_3} \alpha_{3,k_3}^{(i)} \psi_{3,k_3}((\boldsymbol{\beta}_i * \mathbf{a}_1)^*(n), \mathbf{a}_1(n)w_i)$$
$$= \sum_{k_3=1}^{K_3} (\zeta_{3,k_3}^{(i)} * a_1^*)[n - m'_{3,k_3}] \frac{(a_1[n - m'_{3,k_3}])^2}{|a_1[n - m'_{3,k_3}]|^2} |a_1[n - m_{3,k_3}]|^{2p_{3,k_3}}, \tag{2.44}$$

where

$$\zeta_{3,k_3}^{(i)}[n] = \alpha_{3,k_3}^{(i)} |w_i|^{2(p_{3,k_3}-1)} w_i^2 \beta_i^*[n]. \tag{2.45}$$

From examining (2.32), (2.42), (2.44), it is evident that the dual-input PA models can be reformulated into an equivalent SISO model with basis functions that depend only on $\mathbf{a}_1(n)$. Compared to the model without mismatch, we note that there are additional basis functions contributed by $f_2(\cdot)$ and $f_3(\cdot)$. These basis functions can potentially have higher memory or non-linear order than those contributed by $f_1(\cdot)$, and hence the total number of basis function may increase above what is necessary when the PAs are isolated from mismatch and coupling.

We can define the equivalent SISO model as:

$$\bar{f}_{\text{PA}}(\mathbf{a}_1(n); \boldsymbol{\zeta}_1^{(i)}(n), \boldsymbol{\zeta}_2^{(i)}(n); \boldsymbol{\zeta}_3^{(i)}(n)) \triangleq \bar{f}_1(\mathbf{a}_1(n); \boldsymbol{\zeta}_1^{(i)}) + \bar{f}_2(\mathbf{a}_1(n); \boldsymbol{\zeta}_2^{(i)}) + \bar{f}_3(\mathbf{a}_1(n); \boldsymbol{\zeta}_3^{(i)}), \tag{2.46}$$

where

$$\bar{f}_1(\mathbf{a}_1(n); \boldsymbol{\zeta}_1^{(i)}) = \sum_{k_1=1}^{K_1} \zeta_{1,k_1}^{(i)} a_1[n - m'_{1,k_1}] |a[n - m_{1,k_1}]|^{2p_{1,k_1}}, \tag{2.47}$$

$$\bar{f}_2(\mathbf{a}_1(n); \boldsymbol{\zeta}_2^{(i)}) = \sum_{k_2=1}^{K_2} (\zeta_{2,k_2}^{(i)} * a_1)[n - m'_{2,k_2}] |a_1[n - m_{2,k_2}]|^{2p_{2,k_2}}, \tag{2.48}$$

$$\bar{f}_3(\mathbf{a}_1(n); \boldsymbol{\zeta}_3^{(i)}) = \sum_{k_3=1}^{K_3} (\zeta_{3,k_3}^{(i)} * a_1^*)[n - m'_{3,k_3}] \frac{(a_1[n - m'_{3,k_3}])^2}{|a_1[n - m'_{3,k_3}]|^2} |a_1[n - m_{3,k_3}]|^{2p_{3,k_3}}. \tag{2.49}$$

We also can elect to simplify the equivalent SISO model by further pruning some of the basis terms. We prune all basis terms in $\bar{f}_3(\cdot)$ containing the cross-memory product

$$a_1^*[n - q](a_1[n - m'_{3,k_3}])^2, m'_{3,k_3} \neq q \tag{2.50}$$

by setting $\zeta_{3,k_3}^{(i,j)}[n] = 0, n \neq 0$. With this pruning strategy, the remaining basis terms in (2.49) can be written as:

$$\begin{aligned}\bar{f}_3(\mathbf{a}_1(n); \boldsymbol{\zeta}_3^{(i)}) &= \sum_{k_3=1}^{K_3} \zeta_{3,k_3}^{(i)}[0] a_1^*[n - m'_{3,k_3}] \frac{(a_1[n - m'_{3,k_3}])^2}{|a_1[n - m'_{3,k_3}]|^2} |a_1[n - m_{3,k_3}]|^{2p_{3,k_3}} \\ &= \sum_{k_3=1}^{K_3} \zeta_{3,k_3}^{(i)}[0] a_1[n - m'_{3,k_3}] |a_1[n - m_{3,k_3}]|^{2p_{3,k_3}}\end{aligned}$$

All the remaining terms in $\bar{f}_{\text{PA}}(\cdot)$, now contain basis in the same form as the GMP used in the original forward wave basis functions $\psi_{1,k_1}(\cdot)$. Hence, we can then combine the 3 summations over K_1, K_2, and K_3 into a total summation over all K basis terms. It can be shown that the coefficients can be written in the form of output coupling matrices $\mathbf{C}^{(k)} \in \mathbb{C}^{N \times N}$, $k \in \{1, 2, ..., K\}$, which lumps the overall effect of $\boldsymbol{\zeta}_1^{(i)}[n]$, $\boldsymbol{\zeta}_2^{(i)}[n]$, and $\boldsymbol{\zeta}_3^{(i)}[n]$ such that:

$$\bar{f}_{\text{PA}}(\mathbf{a}_1(n); \boldsymbol{\zeta}_1^{(i)}, \boldsymbol{\zeta}_2^{(i)}, \boldsymbol{\zeta}_3^{(i)}) = \sum_{k=1}^{K} \sum_{j=1}^{N} C_{i,j}^{(k)} \alpha_k^{(j)} |w_j|^{2p_k} w_j a_1[n - m'_k] |a_1[n - m_k]|^{2p_k}, \tag{2.51}$$

where $\alpha_k^{(j)}$ corresponds to the k'th basis of the j'th PA's forward envelope model without coupling or mismatch. This output coupling matrix can be seen as an alternative method to view the dominant effect of PA load modulation due to mutual coupling on the forward-traveling envelope.

By substituting $\bar{\mathbf{h}}^{(k)} = \mathbf{h}\mathbf{C}^{(k)}$, the non-ideal user received signal, $\bar{r}[n]$, can be written by modifying (2.6) as:

$$\bar{r}[n] = \sum_{i=1}^{N} \bar{h}_i^{(k)} w_i \sum_{k=1}^{K} \alpha_k^{(i)} |w_i|^{2p_k} |\tilde{x}[n - m_k]|^{2p_k} \tilde{x}[n - m'_k] \tag{2.52}$$

$$= \sum_{k=1}^{K} \gamma_k |\tilde{x}[n - m_k]|^{2p_k} \tilde{x}[n - m'_k], \tag{2.53}$$

$$\gamma_k = \sum_{i=1}^{N} \alpha_k^{(i)} \bar{h}_i^{(k)} w_i \tag{2.54}$$

$$\bar{h}_i^{(k)} = \sum_{j=1}^{N} C_{j,i}^{(k)} h_i = C_{i,i}^{(k)} h_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} C_{j,i}^{(k)} h_j. \tag{2.55}$$

LOS channel coefficients are then substituted for an observation point at (θ_o,ϕ_o) while the array beam is steered to (θ,ϕ) by letting $h_i = h_i^{\mathrm{LOS}}(\theta_o, \phi_o)$ and $w_i = g_i(\theta, \phi) w_i^{\mathrm{MRT}}(\theta, \phi)$. The $g_i(\theta, \phi)$ accounts for the amplitude variation in the phase shifters as the beam is steered. The observed signal in this state, $\bar{r}(\theta_o, \phi_o, \theta, \phi)[n]$ and the SISO PA array coefficients are given by:

$$\bar{r}(\theta_o, \phi_o, \theta, \phi)[n] = \sum_{k=1}^{K} \gamma_k(\theta_o, \phi_o, \theta, \phi) |\tilde{x}[n - m_k]|^{2p_k} \tilde{x}[n - m'_k], \tag{2.56}$$

$$\gamma_k(\theta_o, \phi_o, \theta, \phi) = \gamma_k^{\mathrm{cor}}(\theta_o, \phi_o, \theta, \phi) + \gamma_k^{\mathrm{ucor}}(\theta_o, \phi_o, \theta, \phi) \tag{2.57}$$

$$\gamma_k^{\mathrm{cor}}(\theta_o, \phi_o, \theta, \phi) = \sum_{i=1}^{N} \alpha_k^{(i)} C_{i,i}^{(k)} h_i^{\mathrm{LOS}}(\theta_o, \phi_o) |g_i(\theta, \phi)|^{2p_k} g_i(\theta, \phi) w_i^{\mathrm{MRT}}(\theta, \phi) \tag{2.58}$$

$$\gamma_k^{\mathrm{ucor}}(\theta_o, \phi_o, \theta, \phi) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \alpha_k^{(i)} C_{j,i}^{(k)} h_j^{\mathrm{LOS}}(\theta_o, \phi_o) |g_i(\theta, \phi)|^{2p_k} g_i(\theta, \phi) w_i^{\mathrm{MRT}}(\theta, \phi). \tag{2.59}$$

The non-ideal user-received signal, $\bar{r}(\theta, \phi, \theta, \phi)[n]$ is then given by ((2.56)) with:

$$\gamma_k(\theta, \phi, \theta, \phi) = \gamma_k^{\mathrm{cor}}(\theta, \phi, \theta, \phi) + \gamma_k^{\mathrm{ucor}}(\theta, \phi, \theta, \phi) \tag{2.60}$$

$$\gamma_k^{\mathrm{cor}}(\theta, \phi, \theta, \phi) = \sum_{i=1}^{N} \alpha_k^{(i)} C_{i,i}^{(k)} h_i^{\mathrm{LOS}}(\theta, \phi) |g_i(\theta, \phi)|^{2p_k} g_i(\theta, \phi) w_i^{\mathrm{MRT}}(\theta, \phi) \tag{2.61}$$

$$= \sum_{i=1}^{N} \alpha_k^{(i)} C_{i,i}^{(k)} |g_i(\theta, \phi)|^{2p_k} g_i(\theta, \phi) \tag{2.62}$$

$$\gamma_k^{\mathrm{ucor}}(\theta, \phi, \theta, \phi) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \alpha_k^{(i)} C_{j,i}^{(k)} h_j^{\mathrm{LOS}}(\theta, \phi) |g_i(\theta, \phi)|^{2p_k} g_i(\theta, \phi) w_i^{\mathrm{MRT}}(\theta, \phi). \tag{2.63}$$

Contrary to the ideal scenario without coupling and cross-talk, the user-received signal depends on the beam-steering angle (θ, ϕ). Hence, a new SISO model, $f(\tilde{\mathbf{v}}[n]; \boldsymbol{\gamma}(\theta, \phi, \theta, \phi))$, is potentially needed for every choice of (θ, ϕ). The $\gamma_k(\theta, \phi, \theta, \phi)$ are decomposed into correlated components, $\gamma_k^{\mathrm{cor}}(\theta, \phi, \theta, \phi)$ and uncorrelated components, $\gamma_k^{\mathrm{ucor}}(\theta, \phi, \theta, \phi)$. The number of sets of γ_k necessary to cover a desired range of (θ,ϕ) is expected to depend on both γ_k^{cor} and γ_k^{ucor}. Assuming identical PAs (i.e. $\alpha_k^{(i)} = \alpha_k$), symmetric phase shifter gain variation ($g_i(\theta, \phi) = g_i(-\theta, \phi)$) and symmetric coupling ($C_{j,i}^{(k)} = C_{i,j}^{(k)}, \forall k$), it can be shown that:

$$\bar{r}(\theta, \phi, \theta, \phi) = \bar{r}(-\theta, \phi, -\theta, \phi), \tag{2.64}$$

which implies that under these conditions, the signal steered to and received at (θ,ϕ) is the same as for ($-\theta$,ϕ). This property is useful when considering the total number of SISO model coefficient sets needed to cover the entire steering range of the array.

The SISO model can be extended to accommodate wideband antenna impedance variation and channel responses through increasing the memory depth M of the PA behavioral function. The length of M will depend on the frequency variation of the overall SISO model response. However, by increasing M, the total number of coefficients K is expected to increase. Furthermore, it is expected to be more difficult to control the PA variation and coupling symmetry due to variations in bias, process, layout, and assembly across a wider bandwidth.

2.5 Single-Input Single-Output Digital Predistortion of 64-Element PA Array

To study the applicability of the SISO formulation for DPD, a 64-element array was linearized using the following experimental measurement setup depicted in Fig. 2.8. Following Fig. 2.8, the measurement procedure can be described in 6 steps:

1. A baseband modulated test signal is generated and upconverted to RF carrier.
2. The RF signal is fed into a 64-element array, which includes all RF beamformer circuitry (phase shifters, variable gain attenuators, PAs, and antennas).
3. The array is set to steer its beam on a constant $\phi = 0^\circ$ slice to an angle θ.
4. The array is mounted on a motor, which can rotate the array along the constant $\phi = 0^\circ$ slice.
5. A receiving horn antenna is placed in a fixed position in the array's far-field. The angular position of the of the E-plane of the horn antenna relative to the array's broadside is set using the motor.
6. The output of the horn antenna is fed to a receiver, which downconverts the signal into baseband.

Using the received signals generated in this fashion, a SISO DPD can be trained at a particular beam-steering angle and receiver angular position.

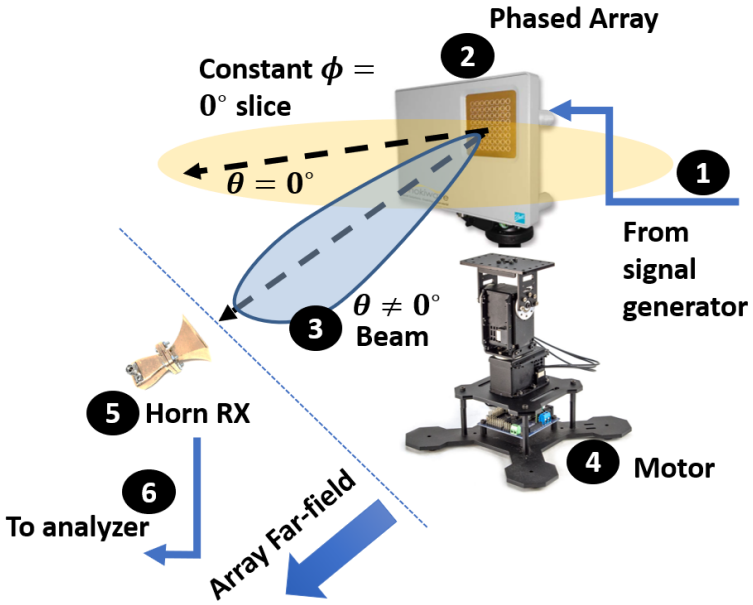

Figure 2.8: Block diagram of the 64-element array measurement configuration.

2.5.1 64-Element Array Measurement Setup

The measurement setup for the 64-element array is shown in Fig. 2.9. A Keysight M8190A Arbitrary Waveform Generator (AWG) is used to generate the baseband In-phase and Quadrature (IQ) signals. A Marki Microwave MMIQ-1037H upconverting mixer translates the baseband signal to a 28 GHz center frequency. The upconverted signal is fed into an Anokiwave AWMF-0129 radio head [5], which contains the RF beamformer, PAs, and antennas within a 64-element array. The AWMF-0129 has an EIRP 1-dB compression point of 50 dBmi. It is attached to a motor which can be adjusted to align the beam to maximize the power at the horn antenna probe. The receiving horn antenna is placed in the far-field of the array and its output is connected to the receiver. The receiver consists of a Marki Microwave MM1-1140H downconverting mixer and DSA91304A 40 Gsps oscilloscope. Independent 800 MHz OFDM signals with a subcarrier spacing of 120 kHz and 9 dB PAPR are used for training and validation the DPD.

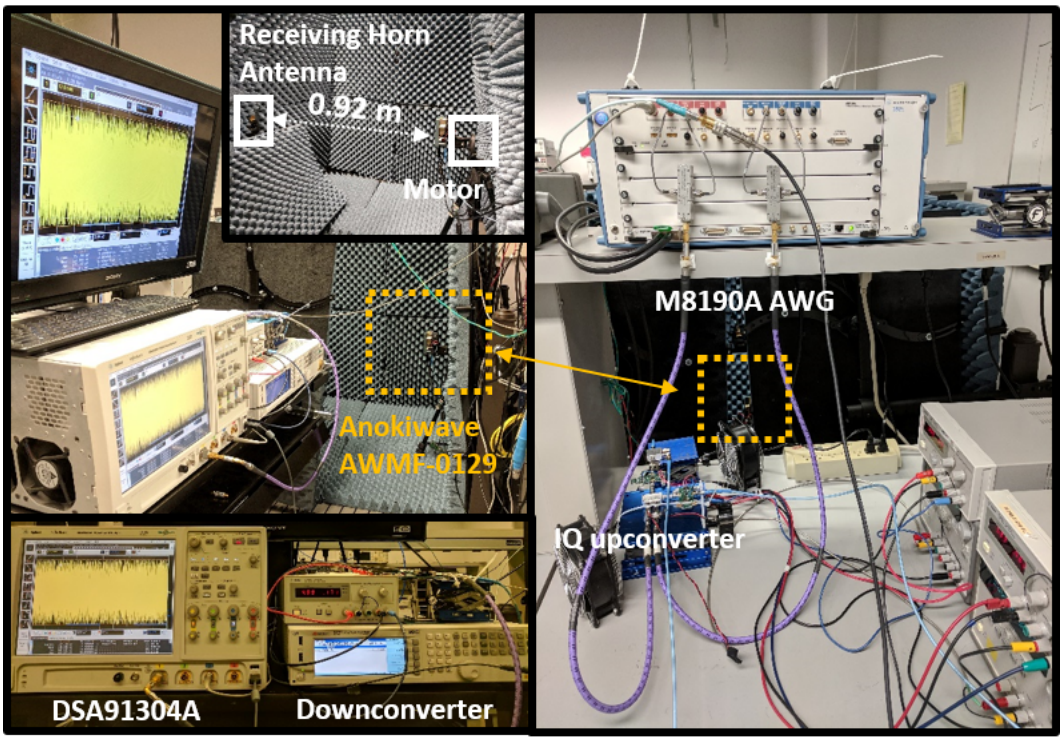

Figure 2.9: Experimental measurement setup for DPD of 64-element Array.

2.5.2 64-Element Array Measurement Results Versus Observation Angle

In this measurement, the array's beam is steered to broadside and motor is set to align the horn antenna to the beam. A SISO DPD function was trained from the horn antenna's received signal using a generalized memory polynomial basis with maximum nonlinearity order $2p_k + 1 = 13$, $M = 4$, and $K = 126$. The corresponding spectrum, AM/AM and AM/PM before and after DPD is shown in Fig. 2.10.

The Adjacent Channel Power Ratio (ACPR), defined as the ratio between the power of an of the adjacent channel to the main channel's power, is measured before and after applying DPD. From the measured results, an ACPR improvement from -34 dB to -49 dB was observed. 2 dB of gain compression and 5 degrees of phase distortion have also been corrected after applying DPD.

The Normalized Mean-Squared Error (NMSE) between an input $\tilde{x}[n]$ and its corre-

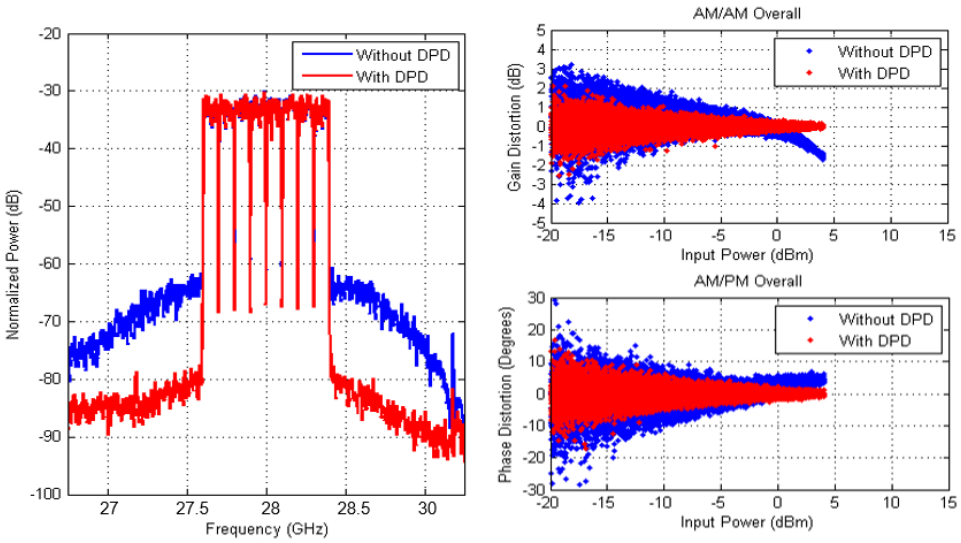

Figure 2.10: Measured spectrum (left), AM/AM and AM/PM (right) of the overall 64-element RF beamforming array system before and after DPD.

sponding time-aligned output $\tilde{y}[n]$ over L samples is defined using the following equation:

$$\text{NMSE} = \frac{\frac{1}{L}\sum_{n=1}^{L}|\tilde{y}[n]-\tilde{x}[n]|^2}{\mathbf{E}[|\tilde{x}[n]|^2]}, \tag{2.65}$$

where $\tilde{x}[n]$ and $\tilde{y}[n]$ are scaled to have the same average power. The NMSE between the common input and Over-the-air (OTA) feedback signal reduced from 5.9% before DPD to 1.9% after DPD.

While keeping the same DPD applied, the motor angle is swept between $\pm 40^\circ$ to change the receiver's observation angle while the beam is fixed to broadside. The power In-band (IB) and OOB are measured. The resulting plot of the IB and OOB power versus observation angle is shown in Fig. 2.11 and is effectively the IB and OOB modulated signal radiation patterns. It is evident from the shape of the IB and OOB radiation pattern in Fig. 2.11 that the SISO DPD only removes nonlinear distortion that is correlated to the IB signal and does not generate spectral content in other directions. The amount of OOB power strictly decreases with DPD and has the most reduction at the direction of maximum combining. These results confirm that no additional OOB power is created in spurious directions.

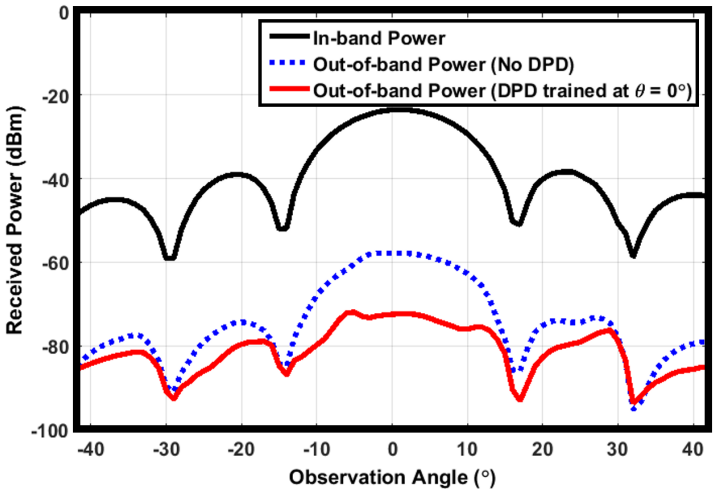

Figure 2.11: Measured modulated signal radiation pattern for IB and OOB power with 800 MHz wide modulated signal.

2.5.3 64-Element Array Measurements Results Versus Steering Angle

In this measurement, the steering angle is then adjusted between within $\theta \in [-60°, 60°]$ while always aligning the relative position of the horn antenna via the motor to receive the maximum power. This alignment is done so that the measurement results for each steering angle can be obtained at the highest SNR. The DPD can subsequently be trained at a specific θ and evaluated for how well the SISO model holds outside the trained angle. Additional DPD coefficients can also be trained at different training angles. The ACPR and NMSE at different training angles are shown in Fig. 2.12 and Fig. 2.13 respectively. To maintain an NMSE below 3% and ACPR below -45 dBc across the steering range, it was found that one training angle was not sufficient. Instead, 5 training angles – at $\theta \in \{0°, \pm 20°, \pm 40°\}$ – were needed. These 5 coefficient sets can be selected using a multiplexed/set-wise approach based on the angular position of the receiver and, as shown in Fig. 2.13, the decision boundaries can be placed based on minimizing the NMSE. This set-wise DPD does not add extra complexity to the DPD basis functions, but requires extra time to train if many coefficients sets are found necessary.

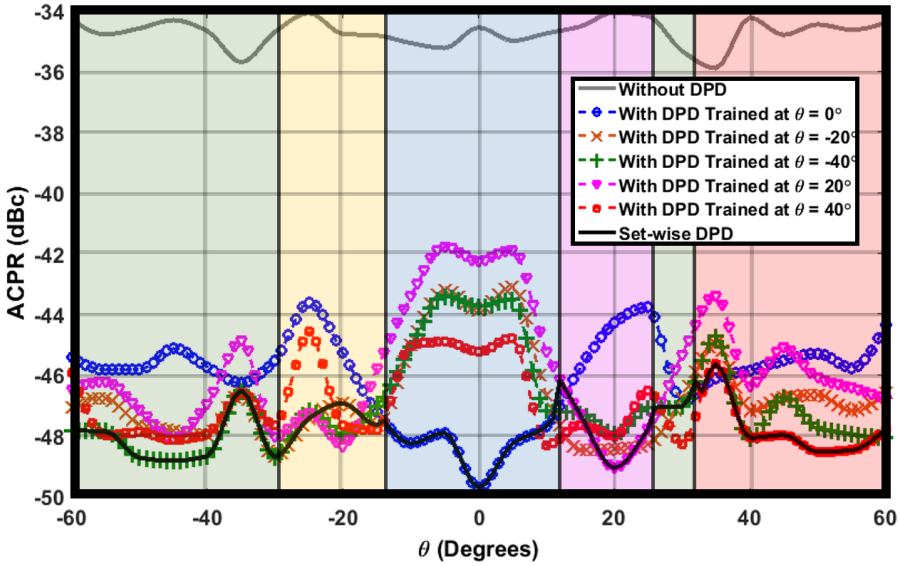

Figure 2.12: Measured ACPR of the 64-element array steered to θ

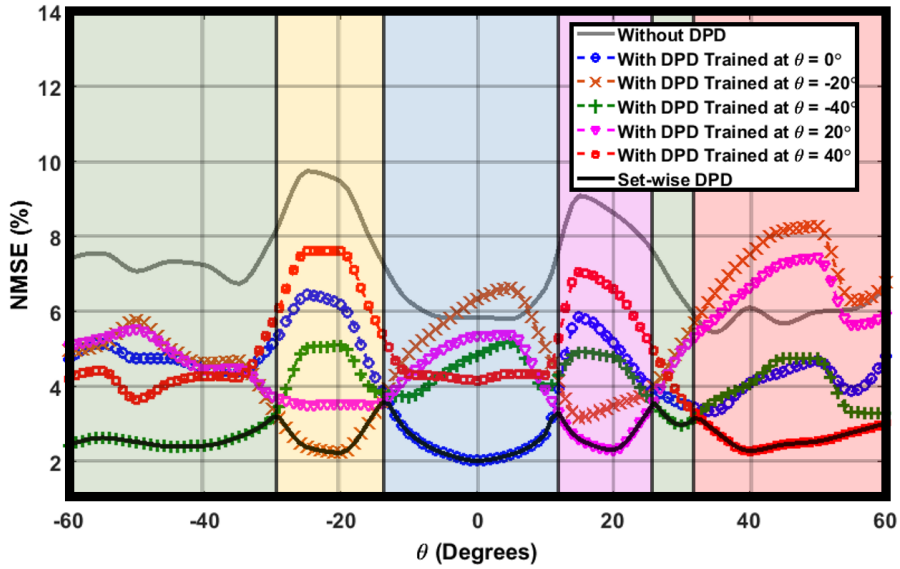

Figure 2.13: Measured NMSE of the 64-element array steered to θ

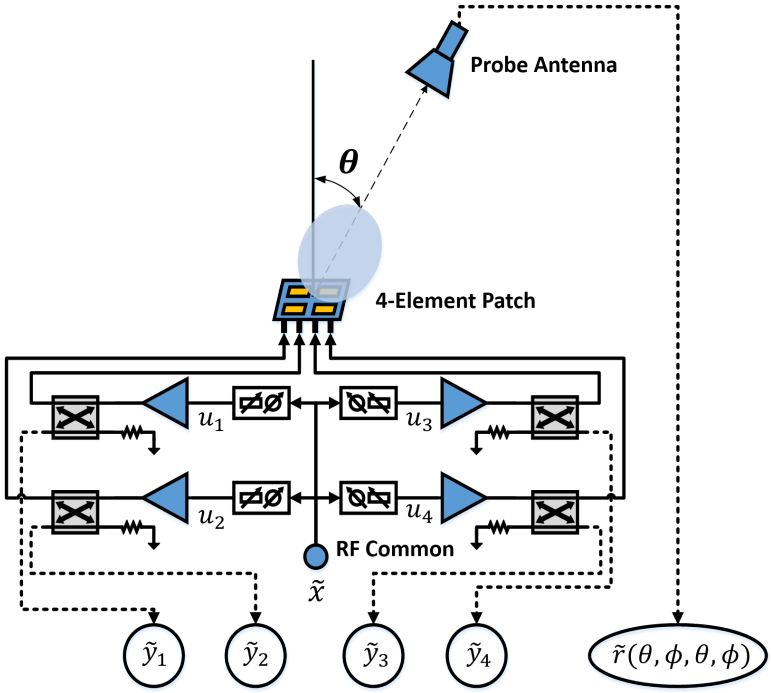

Figure 2.14: Block diagram of the 4-element array measurement configuration

2.6 Single-Input Single-Output Digital Predistortion of 4-Element PA Array

To get a better understanding of the behavior at the output of the individual PAs in the RF beamforming array and methods to reduce the number of SISO coefficient sets, a 4-element array is also considered for SISO DPD, as depicted in Fig. 2.14. According to Fig. 2.14, the array consists of 4 linear patch antennas and are each driven by a PA operating in compression under modulated signal stimulus. The signals to each PA differ only by the phase shift necessary to steer the patch array in the intended direction at θ. The output of each PA is fed through a directional coupler, which allows sampling of the forward wave through the coupled port. As in the 64-element setup, the receiving horn antenna is oriented to always match the direction of maximum antenna gain. The output of the horn antenna is used to train the SISO model, after which its inverse function is applied to generate the predistortion signal.

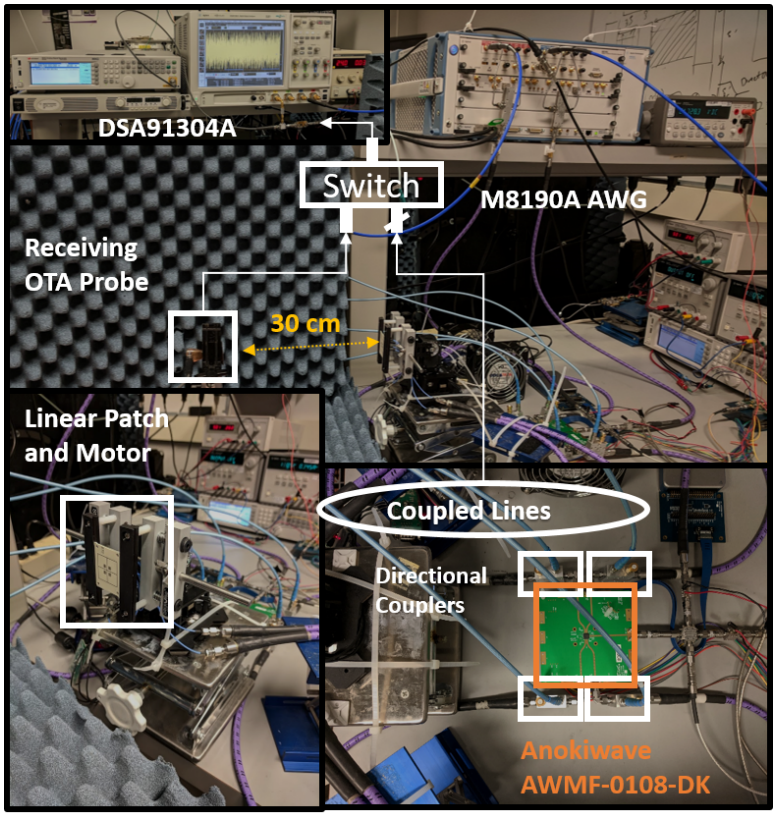

Figure 2.15: Experimental measurement setup of the 4-element configuration

2.6.1 4-Element Array Measurement Setup

The measurement setup for the 4-element array is shown in Fig. 2.15. It includes a Keysight M8190A AWG, which generates the baseband IQ pairs. The baseband signal then feeds into a Marki Microwave MMIQ-1037H upconverting mixer, which upconverts the baseband signal to a 27.5 GHz center frequency. The RF signal is fed into an Anokiwave AWMF-0108 beamforming Integrated Circuit (IC) [38], which provides the PAs, 5-bit digital phase shifters and division circuitry. The PAs have a output 1-dB compression point of 9 dBm, and the output of each of the PAs are fed into a Marki Microwave CA-40 directional coupler where the thru port is attached to the linear patch and the coupled lines are fed into a switch. The 4-element patch array is the same patch as described in Section 2.4.1 and is fixed to a motor which is used to align the beam to the direction of maximum antenna gain. A Marki Microwave MM1-1140H down-converting mixer and DSA91304A 40 Gsps

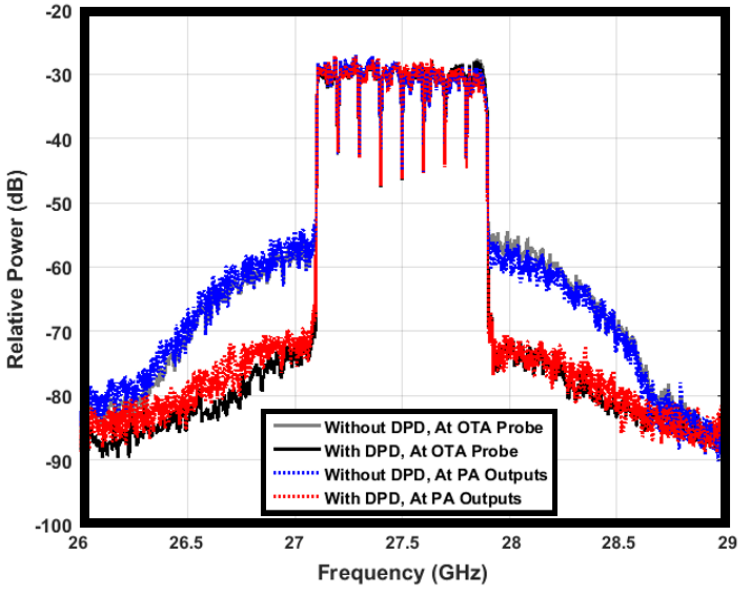

Figure 2.16: Measured spectrum of the received signal and at each of the PA outputs before and after applying DPD.

oscilloscope is used as the receiver, and is able to switch between the output from the receiving horn antenna and each of the coupled PA lines. The beamforming IC is driven by an 800 MHz OFDM signal with a sub-carrier spacing of 120 kHz and 9 dB PAPR.

2.6.2 4-Element Array Measurement Results with Over-The-Air Training Signal

In this measurement, a SISO DPD function with nonlinearity order $2p_k + 1$ up to 13, $M = 13$ and $K = 217$ coefficients is then trained in the broadside direction of the 4-element array with the signal received at the OTA output. The measured spectrum at the PA outputs and at the receiving horn antenna before and after DPD are shown in Fig. 2.16. The spectrum in Fig. 2.16 shows linearization using the SISO DPD is sufficient to simultaneously linearize each PA aside from small variations that exist between the PAs which cannot be linearized by a common DPD function.

The same SISO DPD function trained at broadside is applied while the array is steered between $\pm$ 45° while maintaining alignment of the receiving antenna towards the direction

of maximum radiation of the beam. The ACPR and NMSE are shown versus the steering angle in Fig. 2.17 and Fig. 2.18 respectively.

The AM/AM and AM/PM before and after applying the broadside DPD is measured at the OTA probe output and the 4 PA outputs at the trained angle and at a steering angle of 10°. The AM/AM and AM/PM is shown in Fig. 2.19 and Fig. 2.20 respectively. From Fig. 2.19 (top) it is evident that the PAs have a gain compression of about 2.5 dB before DPD, and this is reflected in the OTA signal. From Fig. 2.20 (top), a phase distortion of up to 10° is observed before DPD. After DPD is applied, the amplitude and phase distortion are fully corrected at the OTA probe when the steering angle kept at broadside.

From the ACPR obtained using SISO DPD trained at $\theta = 0°$ in Fig. 2.17, it is evident that there are peaks for a very small steering angle change to $\theta = 10°$. Changes due to load modulation are not expected to be this drastic with little coupling and small changes in the steering angle. Hence these peaks are primarily attributed to uncalibrated digital phase shifter behavior, which can have up to 1 dB amplitude variation between its phase states as reported in this IC. This is also observed by looking at the AM/AM at the OTA probe and at the individual PA outputs when a steering angle of $\theta = 10°$ is set, shown in Fig. 2.19 (bottom). These phase shifter settings contributed to slightly lower output power, leading to gain expansion at the output of the PAs when the broadside-trained DPD signal is applied.

To address this phase shifter variation, the PAs can be conditioned with initial phase settings that exhibit less amplitude variation from the initial setting across the steering range. This is shown by evaluating two additional DPDs, one trained at $\theta = 10°$ and another trained at $\theta = -10°$. It is evident that these DPDs performed better outside the trained angle by maintaining low ACPR across the entire $\pm 45°$ steering range. It can also be observed that a DPD trained at $\theta = -10°$ achieved lower ACPR than the one trained at $\theta = 10°$, despite being subjected to the same phase shifter settings. This can be caused by any asymmetry in the transmitter paths, of which can vary significantly for small sample sizes and makes it difficult to find the optimal DPD training condition for small arrays.

It is also important to highlight that compared to the ACPR, which remained low across the steering range, we observe from Fig. 2.18 that the NMSE varied much more significantly. This demonstrates that the nonlinear behavior can be less sensitive to a change in steering angle compared to the linear response. It should be therefore possible in low-coupling scenarios to use a single set of SISO DPD coefficients with separate linear filtering functions to compensate for both the in-band and out-of-band distortion across the steering range. This linear filtering can be incorporated in the channel equalization to reduce the overall complexity of the SISO DPD.

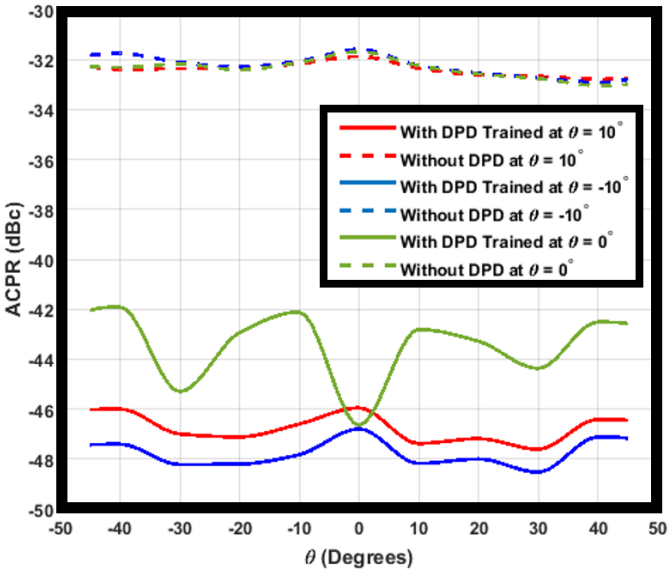

Figure 2.17: Measured ACPR of the 4-element array steered to θ.

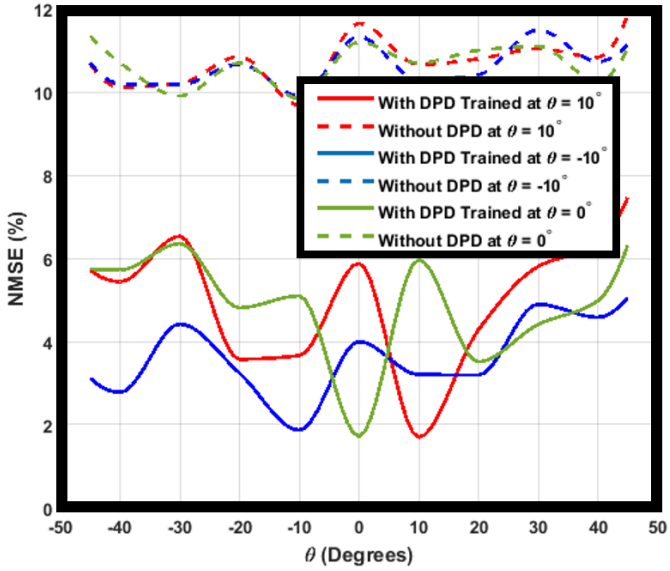

Figure 2.18: Measured NMSE of the 4-element array steered to θ.

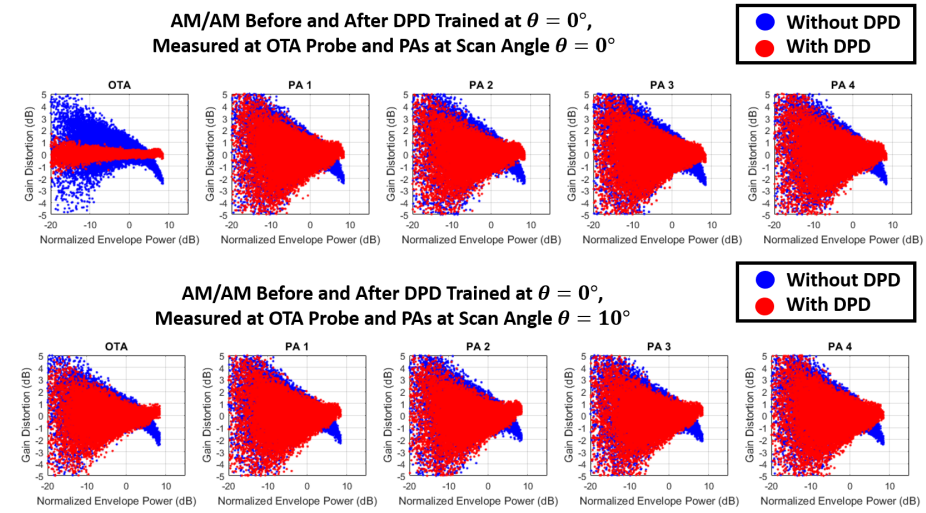

Figure 2.19: Measured AM/AM at the OTA probe and 4 PAs with and without the DPD trained at $\theta = 0°$. The measurement is done with the same steering angle as the training (top) and at a beam-steering angle of $\theta = 10°$ (bottom).

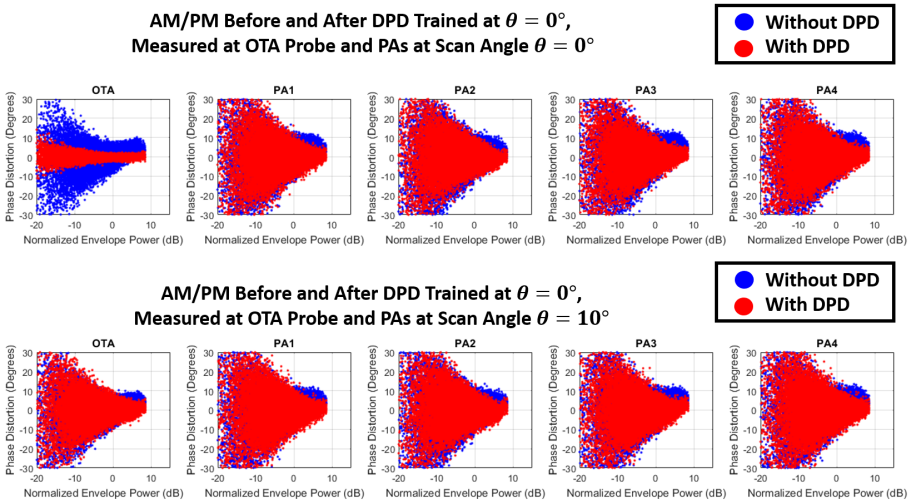

Figure 2.20: Measured AM/PM at the OTA probe and 4 PAs with and without the DPD trained at $\theta = 0°$. The measurement is done with the same steering angle as the training (top) and at a beam-steering angle of $\theta = 10°$ (bottom).

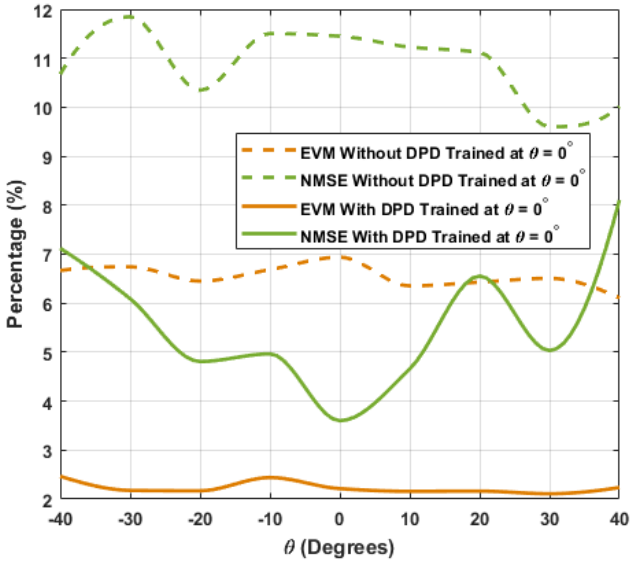

Figure 2.21: Measured NMSE and EVM across the steering angle range before and after applying DPD trained at broadside.

To illustrate this behavior, a separate measurement was performed using a DPD trained at broadside. Note in this measurement, the PAs were pushed to a gain compression of 3 dB, which is why the initial NMSE is higher. In addition to NMSE, the Error Vector Magnitude (EVM) was plotted across the steering angle range and shown in Fig. 2.21. The EVM per OFDM symbol of length L_{FFT} subcarriers is defined as:

$$\frac{\text{EVM}}{\text{symbol}} = \frac{\frac{1}{L_{\text{FFT}}} \sum_{u=1}^{L_{\text{FFT}}} |\tilde{Y}[u]W[u] - \tilde{X}[u]|^2}{\mathbf{E}[|\tilde{X}[u]|^2]}, \tag{2.66}$$

where $\tilde{X}[u]$ and $\tilde{Y}[u]$ are the Discrete Fourier Transform (DFT) of their respective complex baseband envelopes sampled at interval T_c, $T_c = \frac{T_s}{L_{\text{FFT}}}$ and T_s is the OFDM symbol rate. $W[u]$ is a linear equalization filter derived from minimizing the mean-squared error at the pilot subcarriers and interpolating to the remaining subcarriers. The reported EVM is averaged over 10 symbols.

From Fig. 2.21, it is evident that the EVM, which applies a linear equalization filter, is flat across the steering angle range. This is in contrast to the NMSE, which continues to vary significantly with steering angle. This result confirms the applicability of a linear equalization at the receiver to assist the SISO DPD. Given the less-stringent OOB emissions

standards for 5G [18], an equalization-assisted SISO DPD is expected to require very few number of training angles.

2.6.3 4-Element Array Measurement Results with Coupled-line Feedback Signal

In this measurement, the coupled outputs of the PAs are used to train the DPD instead of the output of the OTA probe. This coupled-line based DPD is of interest because it does not require the TOR to be located in the far-field. The PA's coupled outputs are measured one at a time using the switch. The measured outputs are phase-aligned and summed, and the summed signal is then used to train a SISO DPD as normal.

A SISO DPD function with nonlinearity order $2p_k + 1$ up to 13, $M = 13$ and $K = 217$ coefficients is trained with phase shifters set to steer in the broadside direction using the coherent sum of the coupled PA lines. Another SISO DPD with the same model configration is trained using the OTA probe feedback signal at broadside. The two DPDs are then verified at the OTA probe across the steering angle range from $\theta \in [-40°, 40°]$. The measured NMSE and EVM before and after each of the DPDs are applied are shown in Fig. 2.22. The corresponding ACPR across the steering range is shown in Fig. 2.23.

As noticed in these measurement results, when comparing on the basis of NMSE alone, the OTA-based training outperforms the coupled-line training. However, the EVM and ACPR achieved between the training methods are quite similar. The EVM similarity is illustrated in the constellation plot comparing the two DPDs measured at broadside in Fig. 2.24. These results indicate that a DPD trained on the coherent summation of the PA's coupled outputs and assisted with receiver equalization, as is done in the EVM computation, will perform similarly one trained using an OTA feedback signal. It is of note that the output of the directional couplers consists of only the forward traveling waves (or $b_2^{(i)}$'s). The results of this measurement confirms that the forward waves already take into account the effect of load impedance on the PA nonlinearity. Hence, the different between the OTA outputs and coupled outputs would be attributed to the linear frequency response of the antenna, and this is compensated for in the EVM measurement by use of an appropriate equalization filter at the receiver.

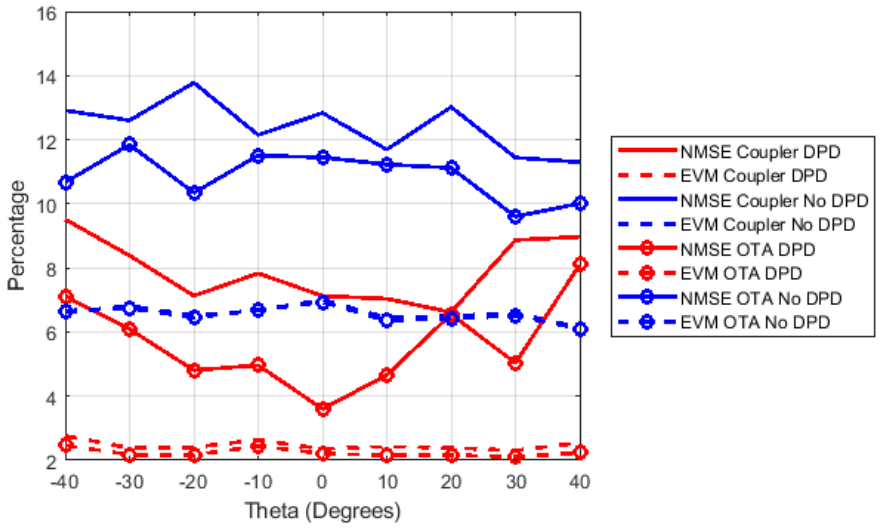

Figure 2.22: Measured NMSE and EVM across the steering angle range before and after applying DPD trained at broadside using the summation of the 4 PA's coupled outputs and OTA.

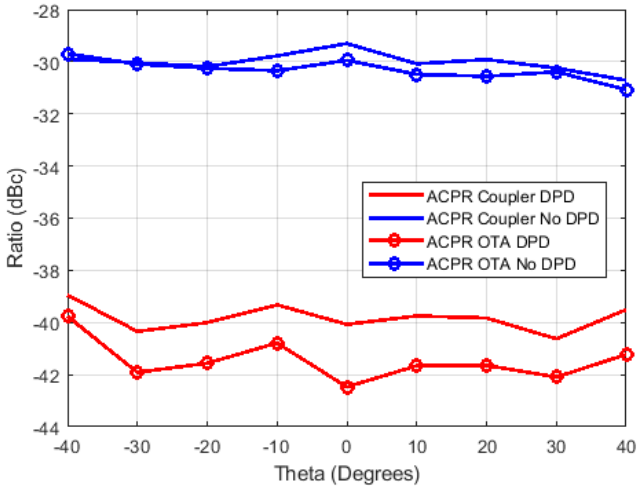

Figure 2.23: Measured ACPR across the steering angle range before and after applying DPD trained at broadside using the summation of the 4 PA's coupled outputs and OTA.

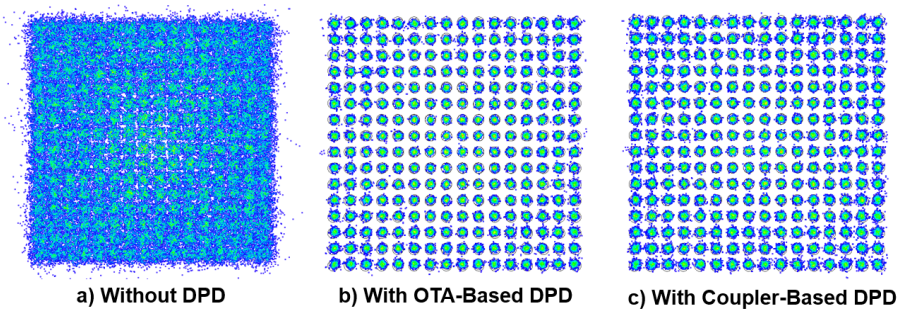

Figure 2.24: Constellation of the received signal before DPD (a), after a DPD trained on the OTA feedback signal (b) and after a DPD trained on the coherent sum of the coupled PA outputs (c).

Chapter 3

Digital Predistortion for Millimeter-wave Multi-User RF Beamforming

The growth in data demands for 5G communications systems necessitates multiple-fold increase in data capacity. As we have seen, one method to achieving higher data capacity is by increasing the available bandwidth through use of mm-wave spectrum. This method requires the use of beamforming to coherently combine an array of antenna's outputs to overcome the higher free-space path loss inherent at mm-wave. Multiple antennas can also be used to achieve spatial diversity. In 3G, 4G and Wireless Local Area Networks (WLAN), multiple-antenna systems were used in multipath-rich environments to allow multiple independent streams of data to be transmitted and received simultaneously. For mm-wave systems, high absorption and high probability for blockage [1] limits achievable SNR and multi-path richness, and hence phase shifter based beamforming is often the optimal linear precoding configuration for achieving high data rate.

However, spatial diversity is not fully lost with RF beamforming. Although each of the antennas signals are now fully correlated with each other, the RF power is directed in a particular direction as a narrow beam. The spatial directivity allows simultaneous transmission of multiple beams to serve multiple users located in different angular directions. This spatial multiplexing through beamforming has the potential to multiply the aggregate data rate at base stations by the number of simultaneous beams created, and is therefore another important consideration in mm-wave phased antenna array systems.

3.1 Subarray-based Multi-user RF Beamforming Arrays

Different methods exist for multi-user beamforming. Digital beamforming [39, 40] and hybrid beamforming [29, 41] architectures offer the ability to create multiple beams from a single array of antennas. However, in RF beamforming, multiple independent beams must be created using separate arrays or sub-divisions of a larger array. These are denoted as subarrays, and an illustration of a typical subarray-based RF beamforming architecture along with a commercial implementation of a 256-element array with 4 64-element subarrays is shown in Fig. 3.1.

As can be seen in Fig. 3.1, the subarray architecture is in essence a scaling-up of the single-user RF beamforming architecture. Each subarray is assigned a separate Digital-to-Analog Converter (DAC) and upconversion circuitry. After the baseband signals are converted to RF, each signal can be individually beamformed using the subarray's phase and gain shifters, PAs, and antennas. Up to V simultaneous users can be supported, where V is the total number of subarrays. Additionally, a linear digital precoder can be used before the DACs. Knowledge of all the baseband signals are available at the digital precoder, and hence the digital precoder can be seen as a $V \times V$ dimension system. Among other uses, the digital precoder can scale, filter, and distribute data streams amongst the subarrays for purposes of optimizing power allocation between the users and mitigating inter-user interference.

3.1.1 Applicability of DPD

Much of the same benefit for applying DPD in single-user RF beamforming applies to the multi-user subarray-based beamforming. This includes improving the power efficiency/nonlinearity trade-off to obtain higher levels of EIRP. In single-user RF beamforming, it was observed in Section 2.5.3 that the unwanted OOB distortion is correlated to the IB signal and therefore beamformed in the same directions as the dominant linear component. As shown Fig. 3.2, when many beams are used, the total OOB power spreads out across the field-of-view of the transmitter, and hence there is a higher probability that the OOB emissions will align in the direction of a victim operating in the adjacent channel. With this in mind, it is perhaps even more critical to apply DPD in such multi-beam systems for OOB distortion mitigation.

In Fig. 3.2, 4 beams, each generated by 64-elements subarrays, are directed to users located at different angles relative to the transmitter. The user locations are indicated by

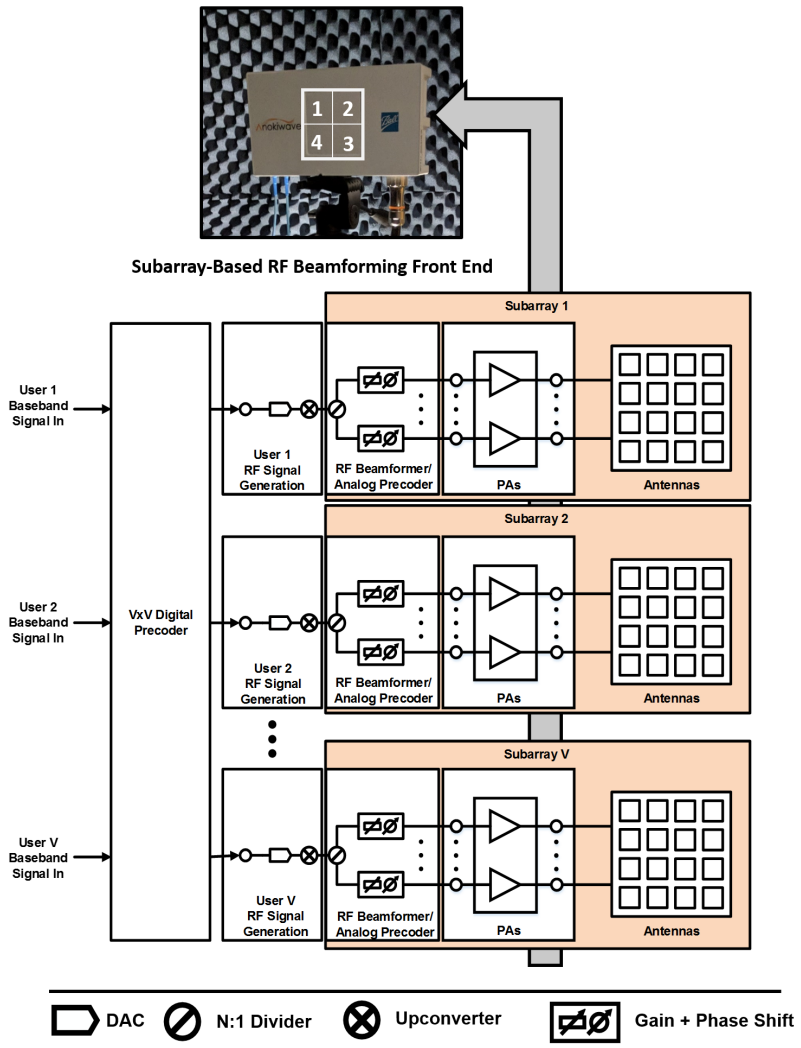

Figure 3.1: Block diagram of a typical subarray-based RF beamforming front-end and a commercial 4x64-Element RF Beamforming Phased Array at 28 GHz [6].

the maxima of the main beams (lobes). However, it is evident from the radiation pattern that for each beam there is spurious spatial radiation outside the main-lobe, called side-lobes. The overlap of these side-lobes onto the main-lobes of other beams causes inter-user interference at each user. The power of these side-lobes relative to the main-lobe vary from

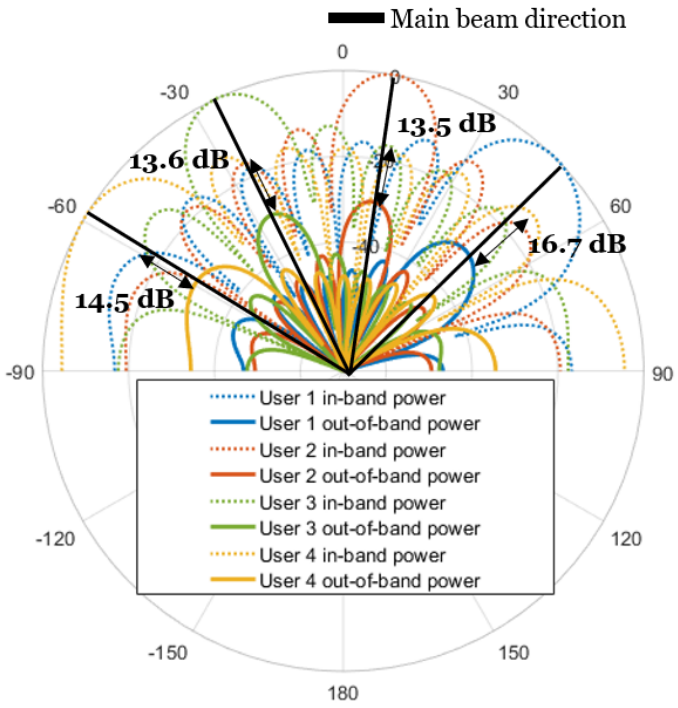

Figure 3.2: Simulated IB and OOB radiation pattern of 4x64-element subarrays. Each subarray is steered to a different angular position with main beam ACPRs of -30 dB. The IB power is shown as dotted and the OOB power is solid. The side-lobes from the IB radiation pattern overlap with the main-lobes of the OOB radiation pattern and have about 15 dB difference.

implementation to implementation depending on the amplitude tapering. In the case when no tapering is applied, the minimum attainable adjacent side-lobe level is -13.5 dB relative to the main-lobe peak [4]. The difference between the IB and OOB power levels is set by the level of compression and subsequently the ACPR. In Fig. 3.2, the OOB radiation pattern corresponding to an ACPR of -30 dB is shown along with the IB radiation pattern. The difference between the IB side-lobes and OOB main-lobes is highlighted along the 4 user directions. It is evident that the IB side-lobes are much higher (around 15 dB) compared to the OOB main-lobe. The existence of the interference from the side-lobes can prove problematic when attempting to characterize the non-linear distortion as is done in DPD training.

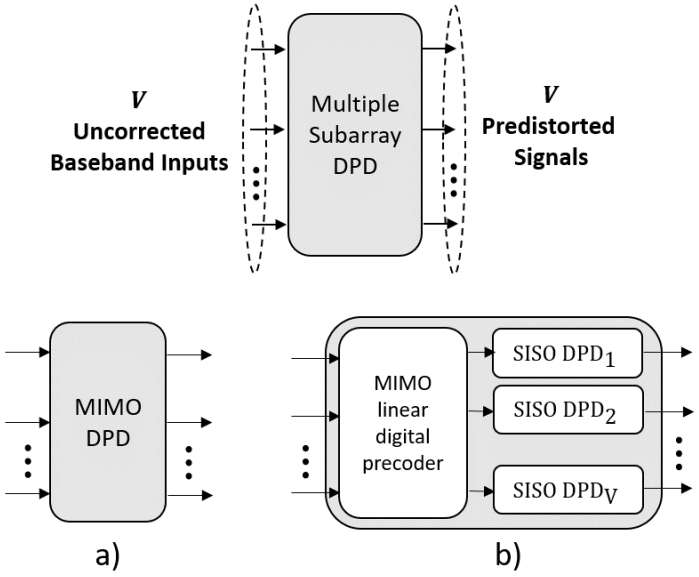

Figure 3.3: Illustration of MIMO DPD (a) versus multiple SISO DPDs assisted by linear digital precoding (b) for the subarray-based RF beamforming array.

3.1.2 Challenges in DPD training complexity

DPD training complexity is comprised of 3 main factors: computational complexity, length of training signal, and the number of TORs or feedback paths.

The computational complexity is associated with the DPD modeling complexity, which can change depending on the number of inputs and outputs, the intensity of the coupling, and the mechanism by which the paths couple to each other. For modeling the subarray-based RF beamforming array in the most general case, a Multiple-Input Multiple-Output (MIMO) DPD can be used which incorporates basis functions from the MIMO Volterra series or pruned versions of it [42]. In these MIMO formulations, the number of required basis functions tends to grow much more rapidly with the number of users than the SISO formulation. Alternatively, one SISO DPD per subarray can be used, as shown in Fig. 3.3. This multiple SISO DPD formulation assumes no significant cross-talk before the input of the PAs and can be assisted by linear digital precoding to compensate for the linear combining that occurs after the PAs. Since the DPD function coefficients are not coupled to each other, less frequent coefficient updates are expected to be necessary.

The length of training signal increases the training complexity as longer training signals require longer capture times, more memory to temporarily store the signal, and more computation to estimate the DPD coefficients. The training signal length is dependent on the several factors. It must be long enough to reflect the statistics of the validation signal (i.e. the CCDFs between the training and validation signals must match closely). The training signal must also be long enough such that noise and other independent sources of error do not bias the coefficient estimation.

Finally, the number of TORs complicates the DPD training setup. Not only would additional hardware be necessary to realize the TORs, but also signals would have to be captured at each of these TORs and then fed back to the central processor for use in DPD training. In closed-loop implementations where the TOR is co-located with the hardware to be linearized, multiple TORs would impose greater cost and implementation challenges. Hence, a DPD training routine requiring only a single TOR or common feedback signal is preferred.

3.1.3 Prior works

In the literature, the TOR architecture in multi-user or multi-antenna systems take various forms. These architectures can be divided into implementations that use a common feedback signal to train multiple DPD functions, and those that use a separate path for each DPD function.

There has been some work that explored the possibility of linearizing multiple or arrays of PAs driven by independent signals with a common feedback receiver. For the TOR, the authors in [43] proposed to use a weighted summation of the output of multiple PAs taken using directional couplers. Each PA is driven with an independent signal and the separate forward models for each PA are jointly estimated. The authors showed through simulation results of memoryless PAs that the common feedback method performs as well as methods that use one feedback signal from one PA at a time, and exceeds the performance of time-sharing method in the presence of crosstalk. In [44], the common feedback signal is obtained through use of one or more OTA TORs. The authors used a forward-model joint-estimation technique with the common feedback signal similar to the one used in [43] to linearize a 4-element digital beamforming array in simulation. These works have high computational complexity due to the need to jointly estimate the forward model coefficients of all the PAs at the same time.

In other works, the DPD functions are estimated one at a time using separate TORs for each DPD function. The authors in [25,27,28] linearize an RF beamforming subarray using

a dedicated TOR for each subarray's DPD function. In [27], a switch-based TOR is used to obtain the output of each PA in the subarray and measurement results are provided for a 2x2 array at 3.5 GHz driven by signals with 10 MHz of modulation bandwidth. [28] proposes an anti-beamforming circuit which uses a coherent summation of the coupled output of the subarray's PAs to train a single DPD for the subarray, which was then validated in simulation using measured models of PAs operating at 2 GHz and driven with 20 MHz digitally modulated signals. Finally, [25] provides simulation results of a memoryless array of PAs using a TOR that uses a downconversion path for each PA. These works do not make use of a common feedback path to train multiple subarrays simultaneously, and hence have higher training setup complexity.

In any of the above TOR configuration, wideband DPD measurement results of mm-wave subarray-based RF beamforming transmitters with multiple users has yet to be reported.

3.1.4 Novelty of this work

In this work, the DPD training complexity in multi-user mm-wave subarray-based RF beamforming transmitters is addressed through application of an OTA TOR assisted with linear interference cancellation. The proposed method uses a single TOR to capture the combined signal from multiple subarrays simultaneously. The SISO DPD functions for the subarrays are then trained in a two-step process. First, an estimate of the linear error, including the inter-user interference, is found and removed from the common feedback signal. Next, the modified feedback signal is used to extract a SISO DPD function for each subarray separately. The two-step process reduces the necessary training signal length and computational overhead compared to other works that require joint-estimation of all DPD coefficients from the combined feedback signal. Using a measurement setup similar to the single-user SISO DPD case (see Section 2.5), the proposed method is validated through linearization of a 2-user RF beamforming system with 2 64-element subarrays operating a 28 GHz and driven by 800 MHz modulated signals. With the two subarrays operating simultaneously, the extracted SISO DPDs are validated across many combinations of the two users' steering angles and an EVM and ACPR improvement of up to 5% and 10 dB respectively was observed across all measured steering angle combinations.

3.2 Nonlinear Modeling of Subarray-based RF Beamforming Arrays

An RF beamforming subarray, shown in Fig. 3.4, can be modeled using a SISO approach similar to the one described in Section 2.2 for a single-user RF beamforming array. Consider subarray $v \in 1, 2, ..., V$ and user $u \in 1, 2, ..., U$. According to the SISO model, if the output of the i'th PA in the v'th subarray can be expressed as a linear combination of Volterra series-based nonlinear basis functions $\psi_k(\cdot)$ such that

$$\tilde{y}_{v,i}[n] = f(\tilde{\mathbf{u}}_v[n]; \boldsymbol{\alpha}_v^{(i)}) = \sum_{k=1}^{K_v} \alpha_{v,k}^{(i)} \psi_{v,k}(\tilde{\mathbf{u}}_v(n)), \tag{3.1}$$

where

$$\tilde{\mathbf{u}}_v(n) = \big(\tilde{u}_v[n], \tilde{u}_v[n-1], \cdots, \tilde{u}_v[n-M]\big) \tag{3.2}$$

is the complex baseband envelope of the input of the PA up to a maximum memory depth M_v and $\boldsymbol{\alpha}_v^{(i)} \in \mathbb{C}^{K_v \times 1}$ is a vector of K_v scalar coefficients, then the overall function that maps the subarray's complex input envelope, $\tilde{\mathbf{x}}_v[n]$ to the corresponding signal received by the u'th observation receiver, $\tilde{r}_{u,v}[n]$ is given by the same basis functions but with different coefficients $\boldsymbol{\gamma}_{u,v} \in \mathbb{C}^{K_v \times 1}$ as

$$\tilde{r}_{u,v}[n] = f(\tilde{\mathbf{x}}_v(n); \boldsymbol{\gamma}_{u,v}) = \sum_{k=1}^{K_v} \gamma_{u,v,k} \psi_{v,k}(\tilde{\mathbf{x}}_v(n)). \tag{3.3}$$

If a GMP [21] is used as the nonlinear basis functions, then

$$\psi_{v,k}(\tilde{\mathbf{x}}_v(n)) = |\tilde{x}_v[n - m_{v,k}]|^{2p_{v,k}} \tilde{x}_v[n - m'_{v,k}], \tag{3.4}$$

where $2p_{v,k}+1$ is the nonlinearity order of the k'th basis and $m_{v,k}$ and $m'_{v,k}$ are lags of the k'th basis. If we also assume that each subarray has N elements, uses a linear precoding vector $\mathbf{w}_v \in \mathbb{C}^{N \times 1}$ and is associated linear channel vector $\mathbf{h}_{u,v} \in \mathbb{C}^{1 \times N}$, then the subarray's coefficients can be written as:

$$\gamma_{u,v,k} = \sum_{i=1}^{N} h_{u,v,i} \alpha_{v,k}^{(i)} |w_{v,i}|^{2p_{v,k}} w_{v,i}. \tag{3.5}$$

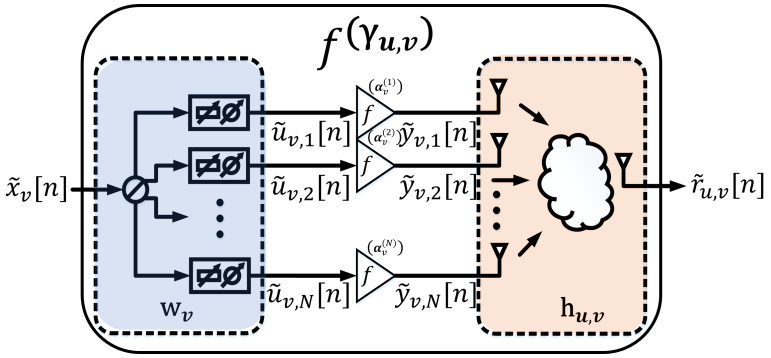

Figure 3.4: Behavioral modeling of an RF beamforming subarray with an equivalent SISO describing function.

If all sub-arrays are transmitting at the same time, the received signal at the u'th receiver, $\tilde{r}_u[n]$ can be written as:

$$\tilde{r}_u[n] = \sum_{v=1}^{V} \tilde{r}_{u,v}[n]. \tag{3.6}$$

For the purpose of describing a multiple subarray and observation receiver system compactly, we can express the SISO coefficients from the V subarrays to U observation receivers in tensor form as:

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,V} \\ \gamma_{2,1} & \gamma_{2,2} & \cdots & \gamma_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{U,1} & \gamma_{U,2} & \cdots & \gamma_{U,V} \end{pmatrix}. \tag{3.7}$$

This can similarly be done for the PA model coefficient vectors, precoding vectors and channel vectors as:

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\alpha}_1^{(1)} & \boldsymbol{\alpha}_1^{(2)} & \cdots & \boldsymbol{\alpha}_1^{(N)} \\ \boldsymbol{\alpha}_2^{(1)} & \boldsymbol{\alpha}_2^{(2)} & \cdots & \boldsymbol{\alpha}_2^{(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}_V^{(1)} & \boldsymbol{\alpha}_V^{(2)} & \cdots & \boldsymbol{\alpha}_V^{(N)} \end{pmatrix}, \tag{3.8}$$

$$\mathbf{w} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_V \end{pmatrix}, \tag{3.9}$$

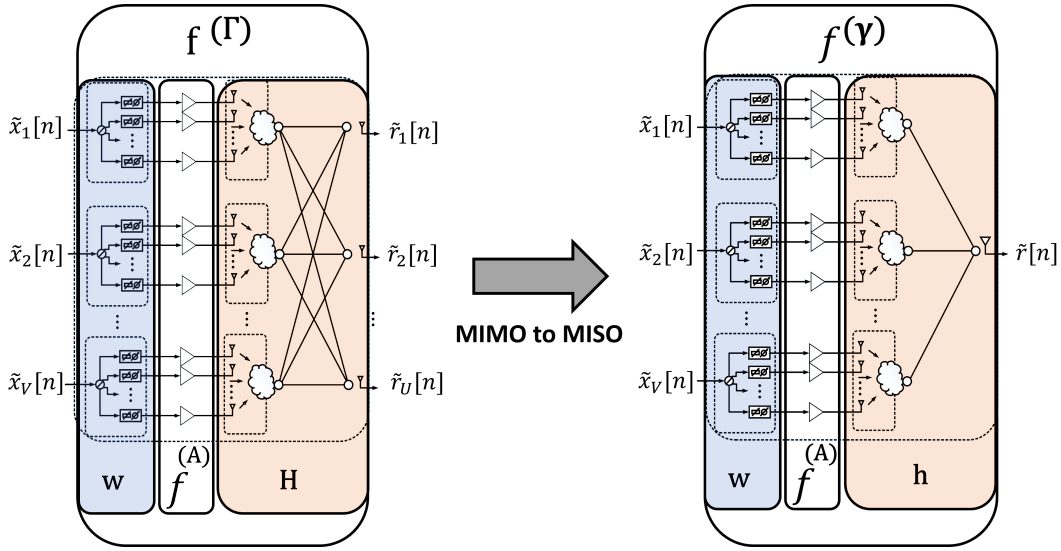

Figure 3.5: Transformation of the general MIMO model (left) for multi-user RF subarray-based beamforming to a MISO model (right).

and

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_{1,1} & \mathbf{h}_{1,2} & \cdots & \mathbf{h}_{1,V} \\ \mathbf{h}_{2,1} & \mathbf{h}_{2,2} & \cdots & \mathbf{h}_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{U,1} & \mathbf{h}_{U,2} & \cdots & \mathbf{h}_{U,V} \end{pmatrix}. \tag{3.10}$$

The overall MIMO function, $\mathbf{f}(\cdot)$ for the subarray system is expressed as

$$\mathbf{f}(\tilde{\mathbf{x}}_1[n], \tilde{\mathbf{x}}_2[n], ..., \tilde{\mathbf{x}}_V[n]; \boldsymbol{\Gamma}) = \begin{pmatrix} \tilde{r}_1[n] \\ \tilde{r}_2[n] \\ \vdots \\ \tilde{r}_U[n] \end{pmatrix} = \begin{pmatrix} \sum_{v=1}^{V} f(\tilde{\mathbf{x}}_v(n); \boldsymbol{\gamma}_{1,v}) \\ \sum_{v=1}^{V} f(\tilde{\mathbf{x}}_v(n); \boldsymbol{\gamma}_{2,v}) \\ \vdots \\ \sum_{v=1}^{V} f(\tilde{\mathbf{x}}_v(n); \boldsymbol{\gamma}_{U,v}) \end{pmatrix}, \tag{3.11}$$

and the overall MIMO system is depicted in Fig. 3.5 (left).

In this analysis, we limit the number of observation receivers to 1 and assume the first receiver is the only receiver. For simplicity of notation, we drop the subscript and thus express the output of the single observation receiver as:

$$\tilde{r}[n] \triangleq \tilde{r}_1[n] = \sum_{v=1}^{V} \tilde{r}_{1,v}[n], \tag{3.12}$$

where

$$\boldsymbol{\gamma}_v \triangleq \boldsymbol{\gamma}_{1,v}. \tag{3.13}$$

We additionally make the following definitions:

$$\boldsymbol{\gamma} \triangleq \begin{pmatrix} \boldsymbol{\gamma}_1 & \boldsymbol{\gamma}_2 & \cdots & \boldsymbol{\gamma}_V \end{pmatrix}, \tag{3.14}$$

$$\mathbf{h}_v \triangleq \mathbf{h}_{1,v}, \tag{3.15}$$

$$\mathbf{h} \triangleq \begin{pmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_V \end{pmatrix}. \tag{3.16}$$

The corresponding Multiple-Input Single-Output (MISO) system can be written as

$$f(\tilde{\mathbf{x}}_1(n), \tilde{\mathbf{x}}_2(n), ..., \tilde{\mathbf{x}}_V(n); \boldsymbol{\gamma}) \triangleq \tilde{r}[n] = \sum_{v=1}^{V} f(\tilde{\mathbf{x}}_v(n); \boldsymbol{\gamma}_v), \tag{3.17}$$

and the conversion from MIMO to MISO is summarized in Fig. 3.5 (right).

3.3 Interference Cancellation For SISO DPD Training

Next we consider the task of finding the SISO modeling coefficients for the v'th subarray, $\boldsymbol{\gamma}_v$. We first examine the case without inter-user interference. This is equivalent to performing the estimation one subarray at a time, with only one subarray enabled in each estimation. The least-squares estimate for the SISO coefficients of the v'th subarray for a length $D+M_v$ of known $\tilde{x}_v[n]$ samples and length D of observed output samples from $\tilde{r}_{1,v}[n]$, $\hat{\boldsymbol{\gamma}}_v$, is found by solving:

$$\hat{\boldsymbol{\gamma}}_v = (\boldsymbol{\Psi}_v^{\mathrm{H}} \boldsymbol{\Psi}_v)^{-1} \boldsymbol{\Psi}_v^{\mathrm{H}} \tilde{\mathbf{r}}_v, \tag{3.18}$$

where

$$\boldsymbol{\Psi}_v = \begin{pmatrix} \boldsymbol{\psi}_v(\tilde{\mathbf{x}}_v(n)) \\ \boldsymbol{\psi}_v(\tilde{\mathbf{x}}_v(n-1)) \\ \vdots \\ \boldsymbol{\psi}_v(\tilde{\mathbf{x}}_v(n-D+1)) \end{pmatrix}, \tag{3.19}$$

$$\boldsymbol{\psi}_v(\tilde{\mathbf{x}}_v(n)) = \begin{pmatrix} \psi_1(\tilde{\mathbf{x}}_v(n)), \psi_2(\tilde{\mathbf{x}}_v(n)), \cdots, \psi_{K_v}(\tilde{\mathbf{x}}_v(n)) \end{pmatrix}, \tag{3.20}$$

$$\tilde{\mathbf{r}}_v = \begin{pmatrix} \tilde{r}_{1,v}[n] \\ \tilde{r}_{1,v}[n-1] \\ \vdots \\ \tilde{r}_{1,v}[n-D+1] \end{pmatrix}, \tag{3.21}$$

and $\mathbf{\Psi}^{\mathrm{H}}$ denotes the Hermitian transpose of matrix $\mathbf{\Psi}$.

In contrast, if all subarrays are enabled, the estimate with interference, $\hat{\boldsymbol{\gamma}}'_v$, is expressed as:

$$\hat{\boldsymbol{\gamma}}'_v = (\mathbf{\Psi}_v^{\mathrm{H}}\mathbf{\Psi}_v)^{-1}\mathbf{\Psi}_v^{\mathrm{H}}\tilde{\mathbf{r}} = \hat{\boldsymbol{\gamma}}_v + \bar{\boldsymbol{\gamma}}_v, \tag{3.22}$$

where

$$\tilde{\mathbf{r}} = \begin{pmatrix} \tilde{r}[n] \\ \tilde{r}[n-1] \\ \vdots \\ \tilde{r}[n-D+1] \end{pmatrix}, \tag{3.23}$$

$$\bar{\boldsymbol{\gamma}}_v = (\mathbf{\Psi}_v^{\mathrm{H}}\mathbf{\Psi}_v)^{-1}\mathbf{\Psi}_v^{\mathrm{H}} \sum_{\substack{u=1 \\ u\neq v}}^{V} \tilde{\mathbf{r}}_u. \tag{3.24}$$

As is evident from (3.22), the estimate with interference is the estimate without interference together with a bias term $\bar{\boldsymbol{\gamma}}_v$. If the $\tilde{x}_v[n]$ are generated from independent zero-mean random processes, then given a sufficient number of samples of $\tilde{r}[n]$, the least-squares estimate for $\bar{\boldsymbol{\gamma}}_v$ would be all zeros. From this, we note that the use of longer signals can get us to the desired SISO coefficients, and this length depends on the magnitude of the interference.

In the MISO case where these interference terms are strong, prohibitively long signals may be required to extract the SISO DPD coefficients. To keep training lengths short, we propose to find an estimate of the interference and subtract it from the feedback signal before it is used for DPD training. Since the PAs are assumed weakly non-linear, linear interference will dominate over nonlinear interference, and thus we consider an interference estimate using only linear basis. The coefficients of linear interference estimate, $\hat{\bar{\boldsymbol{\gamma}}}$ is found by solving

$$\hat{\bar{\boldsymbol{\gamma}}} = (\mathbf{\Phi}^{\mathrm{H}}\mathbf{\Phi})^{-1}\mathbf{\Phi}^{\mathrm{H}}\tilde{\mathbf{r}}, \tag{3.25}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \tilde{\mathbf{x}}_1(n) & \tilde{\mathbf{x}}_2(n) & \cdots & \tilde{\mathbf{x}}_V(n) \\ \tilde{\mathbf{x}}_1(n-1) & \tilde{\mathbf{x}}_2(n-1) & \cdots & \tilde{\mathbf{x}}_V(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{x}}_1(n-L+1) & \tilde{\mathbf{x}}_2(n-L+1) & \cdots & \tilde{\mathbf{x}}_V(n-L+1) \end{pmatrix}, \tag{3.26}$$

and L is the length of the filter to be estimated. With $\mathbf{\Phi}\hat{\bar{\boldsymbol{\gamma}}}$, not only do we have an estimate of the interference, but also an estimate of the linear distortion for each subarray. This is

helpful as now only the non-linear distortion remains to be estimated. The feedback signal with mainly non-linear distortion can be obtained by calculating

$$\bar{r}[n] = \tilde{r}[n] - \mathbf{\Phi}\hat{\tilde{\boldsymbol{\gamma}}} \tag{3.27}$$

and then $\bar{r}[n]$ can be used to train each of the subarray's SISO DPDs separately using conventional inverse modeling techniques that assume an additive non-linear distortion model.

3.4 Digital Predistortion of 2x64-Element RF Beamforming Array

The proposed training method is summarized in 5 steps shown in Fig. 3.6 and described as follows.

1. Each of the V subarrays are fed an independent training signal $\tilde{x}_1[n], \tilde{x}_1[n], ..., \tilde{x}_V[n]$.
2. All V subarrays are set to steer their beams to the same angular position θ_{tr}.
3. The single TOR is also positioned to θ_{tr} to maximize the received power from all of the subarrays.
4. The TOR captures the combined feedback signal $\tilde{r}[n]$. Using the feedback signal, an estimate of the linear distortion is computed and subtracted from the original feedback signal to produce the modified feedback signal $\bar{r}[n]$.
5. Each individual subarray's SISO DPD coefficients are estimated one subarray at a time using $\bar{r}[n]$. The SISO DPD engines are then updated with newly computed coefficients.

These steps can be repeated as required for other θ_{tr} to train additional sets of SISO DPD coefficients. The additional sets of coefficients can be used to construct a set-wise DPD that covers the entire steering range of the subarrays. It is to note that the estimate for linear interference needs to be calculated only once per θ_{tr}.

The same training steps are depicted using a block diagram of the experimental measurement setup shown in Fig. 3.7. The independent signals are sent to 2 64-element subarrays which sit on-top of a motor and are set to beamform the signals on a constant

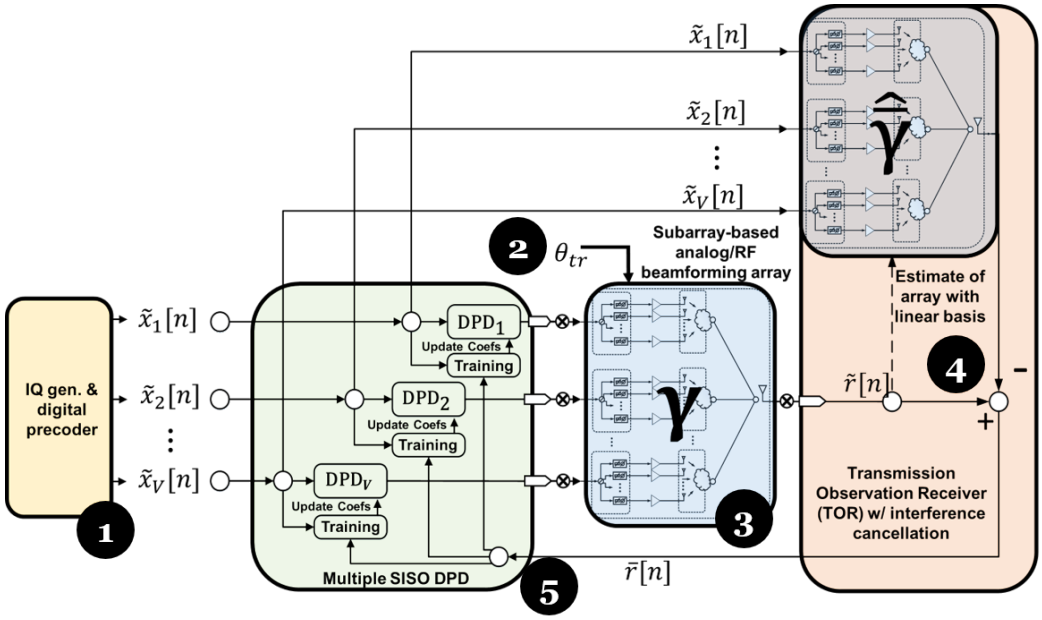

Figure 3.6: Signal flow for DPD training using interference cancellation feedback in subarray-based RF beamforming arrays.

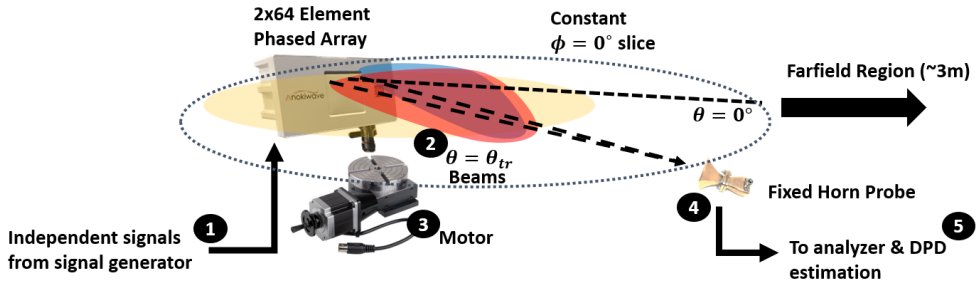

Figure 3.7: Block diagram of the 2x64-element array measurement configuration.

$\phi = 0^\circ$ slice to θ_{tr}. Those signals are received by a fixed probing horn antenna in the array's far-field whose relative position to the subarrays is set by the motor. The output of the horn antenna is then sent to the analyzer where the interference cancellation and SISO DPD estimation steps are performed.

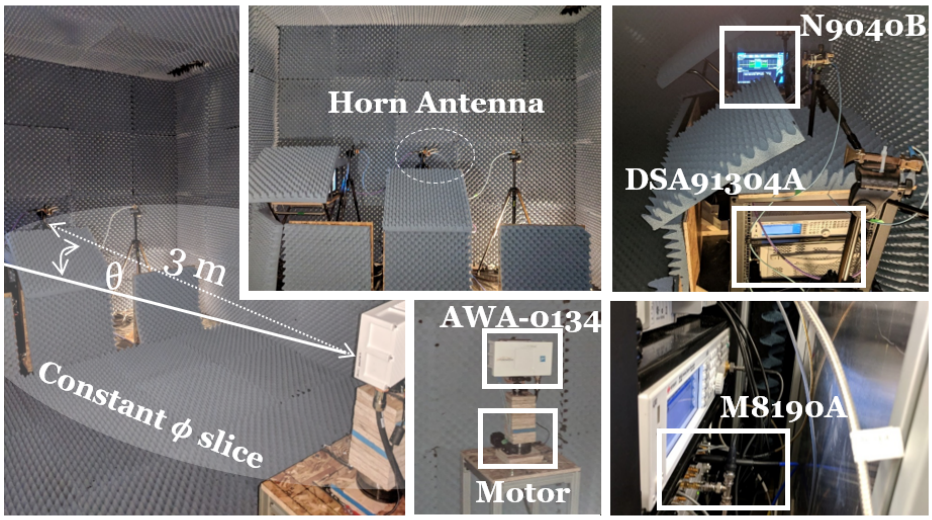

Figure 3.8: Experimental measurement setup for DPD of 2x64-element array.

3.4.1 2x64-Element Array Measurement Setup

The measurement setup for the 64-element array is shown in Fig. 3.8. It includes a Keysight M8190A AWG, where each of its two outputs are used to generate an 800 MHz OFDM signal at a 2.6 GHz Intermediate Frequency (IF). The OFDM signals have 9 dB of PAPR, with each subcarrier spaced 120 kHz apart and 256-QAM modulated. The two signals are then upconverted using a Marki MMIQ-1037H IQ mixer in single-sideband configuration from IF to a 28 GHz RF center frequency. The RF signals are fed into two 64-element subarrays in an Anokiwave AWA-0134 256-element array (shown as 1 and 2 in Fig. 3.1). Each of the subarrays have 48 dBmi EIRP at the 1-dB compression point. A receiving horn antenna is placed in the far-field of the array. The output from the horn antenna is equally split into two outputs, where one output is fed into a Keysight N9040B UXA in spectrum analyzer mode and the other is downconverted using a Marki MM1-1140H mixer to a receiver IF of 1.3205 GHz. The received signal is sampled at IF using a DSA91304A 40 Gsps oscilloscope. All the measurements were conducted within a fully shielded chamber.

3.4.2 2x64-Element Array Measurement Results

A training signal of length 3.75 μs and sampled at 4 Gsps was used to extract the SISO DPD coefficients for both subarrays at $\theta_{tr} = 0°$. The extracted SISO DPDs use a pruned version of the GMP with a maximum nonlinearity order of 7, maximum memory depth of 10, and a total of 35 coefficients per subarray. 60 memory taps (15 ns) per subarray were used in the linear interference cancellation step. The constellation of the verification signal, which is 30 μs or 3 OFDM symbols long, is shown for both users after interference cancellation in Fig. 3.9 with and without DPD.

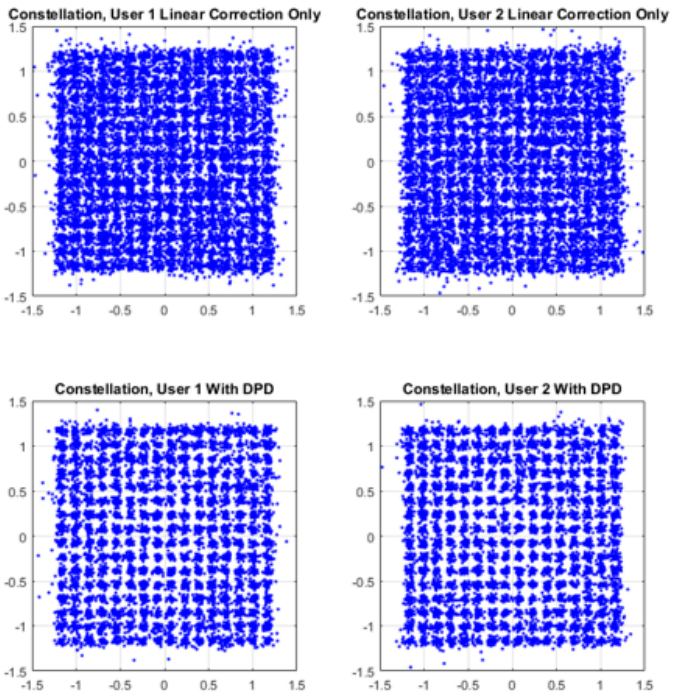

Figure 3.9: Measured constellation at broadside for user 1 (left) and user 2 (right) after linear interference cancellation only (top) and after interference cancellation with DPD (bottom).

Next the same SISO DPDs trained at $\theta_{tr} = 0°$ are verified across all combinations of subarray 1's steering angle ($\theta_1 \in [-60°, 60°]$) and subarray 2's steering angle ($\theta_2 \in [-60°, 60°]$) in steps of 15°. For each (θ_1,θ_2) pair, the motor first positions the horn antenna to maximize the received power at θ_1 where user 1's signal is captured, and then at θ_2 to capture the user 2's signal. The EVM and ACPR are reported for both users in Fig. 3.10

and Fig. 3.11 respectively.

As can be seen in Fig. 3.10 and Fig. 3.11, the EVM and ACPR improved by 5% and 10 dB in the trained directions. However, the trained DPD did not generalize well outside of $\theta_1 = 0°$ and $\theta_2 = 0°$. Hence, additional sets of SISO DPD coefficients were trained using $\theta_{tr} \in [-60, 60]$ in 15° steps. It was found that only 2 distinct cases – at broadside and outside of broadside – existed; all trained angles outside of $\theta_{tr} = 0°$ generalized well to all other angles except broadside. As an example, the EVM and ACPR of the $\theta_{tr} = 15°$ case is shown for all measured combinations of θ_1 and θ_2 in Fig. 3.12 and Fig. 3.13 respectively.

The AM/AM and AM/PM of both the users' signals are shown before and after applying the SISO DPDs trained at $\theta_{tr} = 0°$ in Fig. 3.14. The AM/AM plots are shown on the top, while the AM/PM plots are shown on the bottom. The left plots correspond to user 1 while the right plots correspond to user 2. Each line represents the measured AM/AM or AM/PM at a single (θ_1,θ_2) pair, and the collection of all lines represent all measured combinations of θ_1 and θ_2 across the steering range. It is evident from Fig. 3.14 that a gain compression up to 2.5 dB is experienced at broadside and this compression is higher than at the other steering angles. Hence, while the SISO DPDs trained at these angles performed well when validated at the trained angle, they end up over-correcting at all other angles, which results in gain expansion. Similarly, when examining Fig. 3.15 which shows the AM/AM and AM/PM plots of the users' received signals before and after applying SISO DPDs trained at $\theta_{tr} = 15°$, all angles aside from those at broadside are corrected while the broadside case remains under-corrected, resulting in residual gain compression. The cause of this effect is likely attributed to the uncalibrated phase-shifters, which is shown in Fig. 2.7 to have up to 0.375 dB of output power variation.

Due to these distinct two cases, we can multiplex between the $\theta_{tr} = 0°$ and $\theta_{tr} = 15°$ SISO DPD coefficient sets using a simple strategy: apply the $\theta_{tr} = 0°$ at broadside and $\theta_{tr} = 15°$ coefficients outside broadside. Using this strategy, the achieved levels of EVM and ACPR across all measured combinations of steering angles are shown in Fig. 3.16 and Fig. 3.17 respectively. As is evident from the figures, the set-wise strategy ensured EVM below 3% and ACPR below -40 dB for all measured steering angle combinations.

Lastly, the achieved AM/AM and AM/PM using the set-wise strategy are plotted in Fig. 3.18. From the plots, gain compression ranging from 2.5 dB to 1 dB are corrected at both users after DPD. Phase distortion ranging from 10 degrees to 5 degrees is also compensated. The achieved spectrum at 28 GHz before and after DPD is shown in Fig. 3.19 for user 1 (top) and user 2 (bottom) with all steering angle combinations overlaid. The spectrum plots confirm that the OOB power has been reduced to levels close to the noise floor of the measurement.

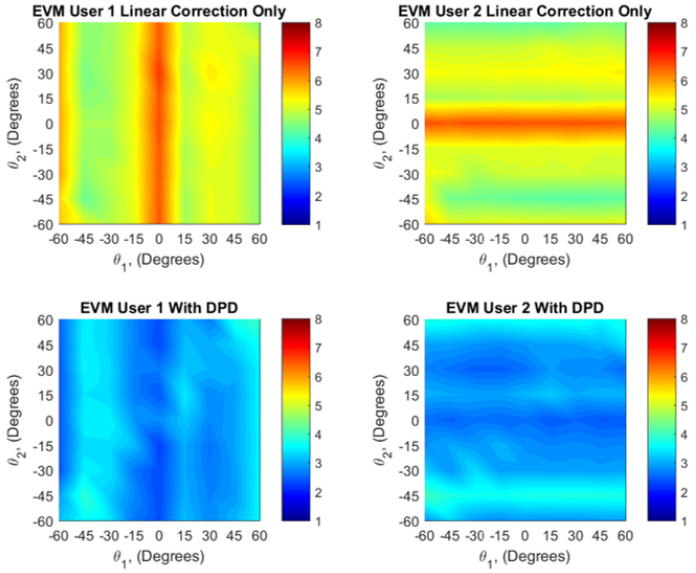

Figure 3.10: Measured EVM (%) at user 1 (left) and 2 (right) for combinations of steering angles of subarray 1 (θ_1) and 2 (θ_2) w/ and w/o applying SISO DPDs trained at $\theta_{tr} = 0°$.

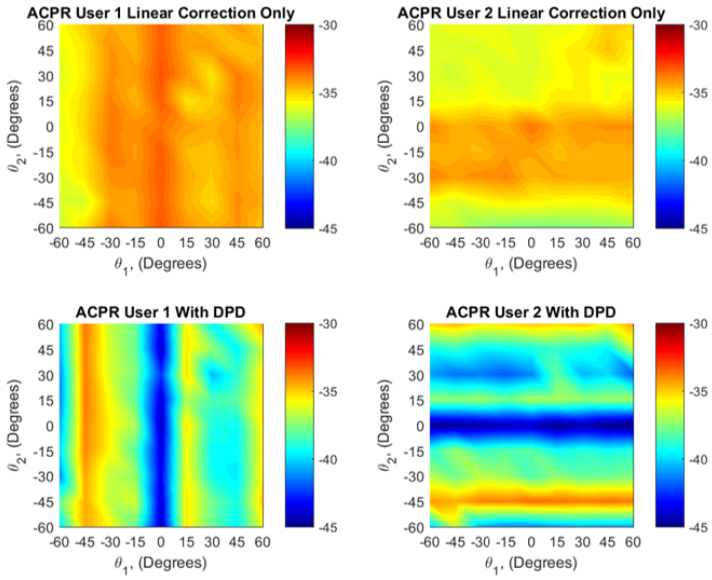

Figure 3.11: Measured ACPR (dB) at user 1 (left) and 2 (right) for combinations of steering angles of subarray 1 (θ_1) and 2 (θ_2) w/ and w/o applying SISO DPDs trained at $\theta_{tr} = 0°$.

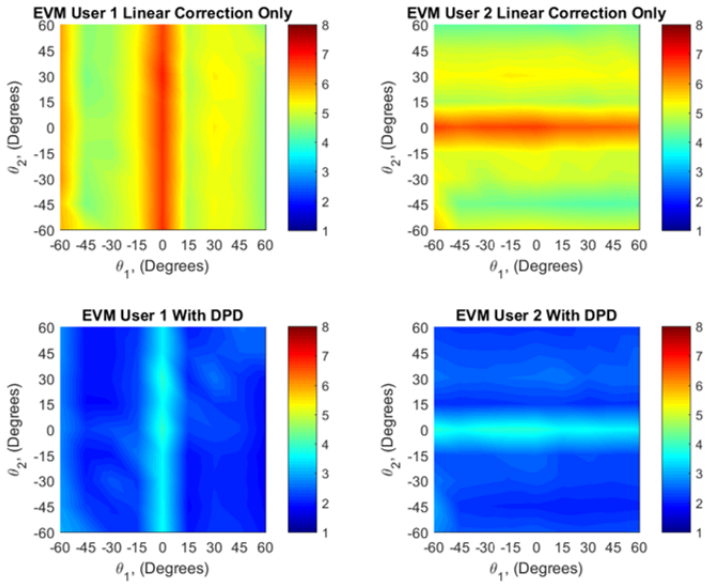

Figure 3.12: Measured EVM (%) at user 1 (left) and 2 (right) for combinations of steering angles of subarray 1 (θ_1) and 2 (θ_2) w/ and w/o applying SISO DPDs trained at $\theta_{tr} = 15°$.

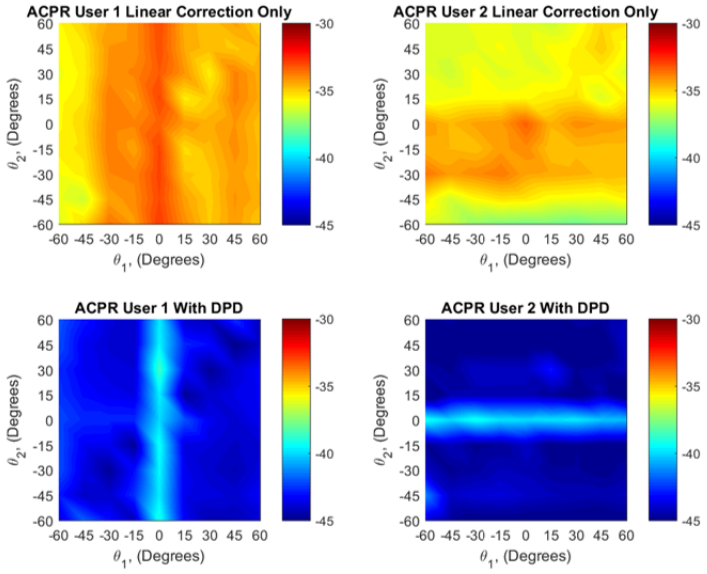

Figure 3.13: Measured ACPR (dB) at user 1 (left) and 2 (right) for combinations of steering angles of subarray 1 (θ_1) and 2 (θ_2) w/ and w/o applying SISO DPDs trained at $\theta_{tr} = 15°$.

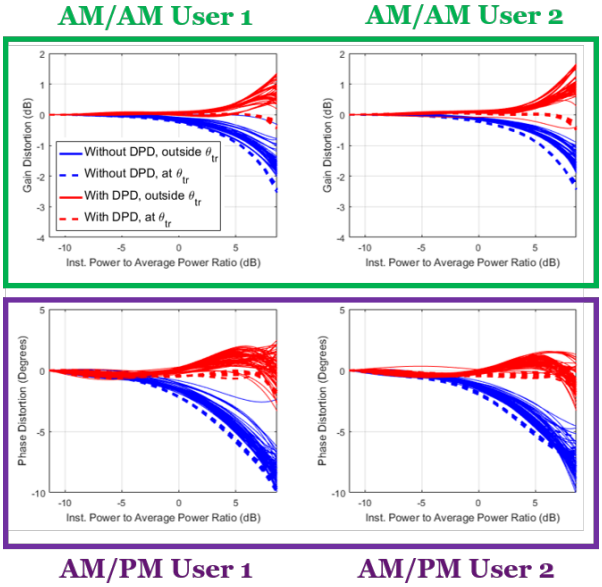

Figure 3.14: AM/AM (top) and AM/PM (bottom) for user 1 (left) and user 2 (right) across all measured steering angles w/ and w/o applying SISO DPDs trained at $\theta_{tr} = 0°$.

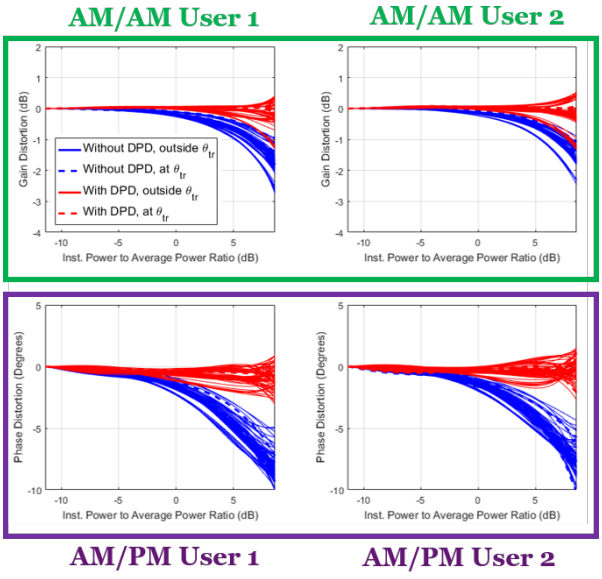

Figure 3.15: AM/AM (top) and AM/PM (bottom) for user 1 (left) and user 2 (right) across all measured steering angles w/ and w/o applying SISO DPDs trained at $\theta_{tr} = 15°$.

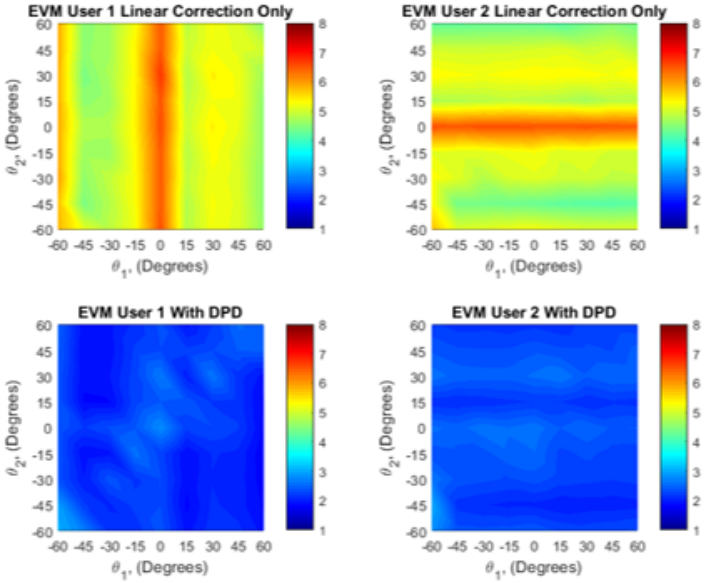

Figure 3.16: Measured EVM (%) at user 1 (left) and 2 (right) for combinations of steering angles of subarray 1 (θ_1) and 2 (θ_2) w/ and w/o set-wise $\theta_{tr} \in \{0°, 15°\}$ SISO DPDs.

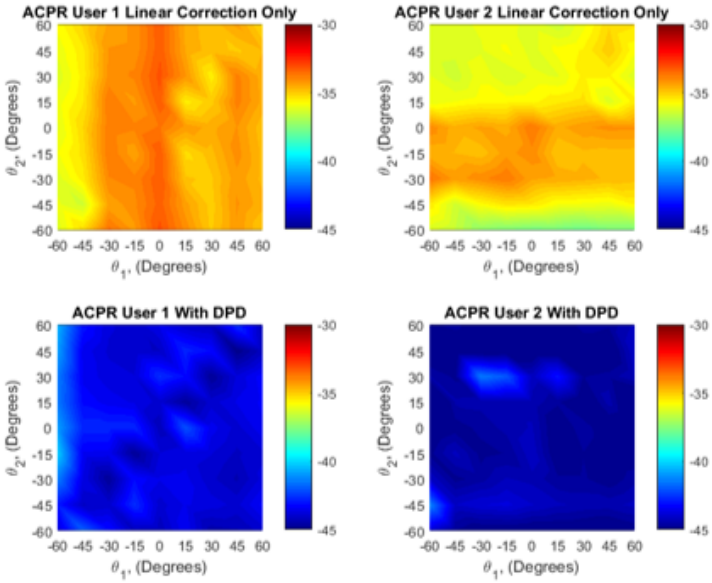

Figure 3.17: Measured ACPR (dB) at user 1 (left) and 2 (right) for combinations of steering angles of subarray 1 (θ_1) and 2 (θ_2) w/ and w/o set-wise $\theta_{tr} \in \{0°, 15°\}$ SISO DPDs.

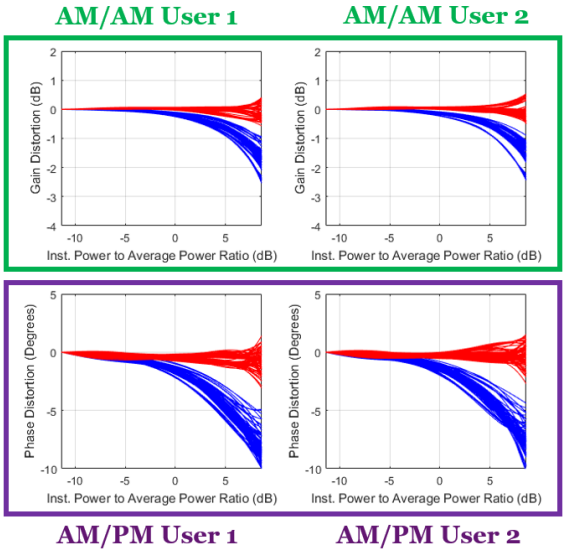

Figure 3.18: AM/AM (top) and AM/PM (bottom) for user 1 (left) and user 2 (right) across all measured steering angles w/ and w/o applying set-wise $\theta_{tr} \in \{0°, 15°\}$ SISO DPDs.

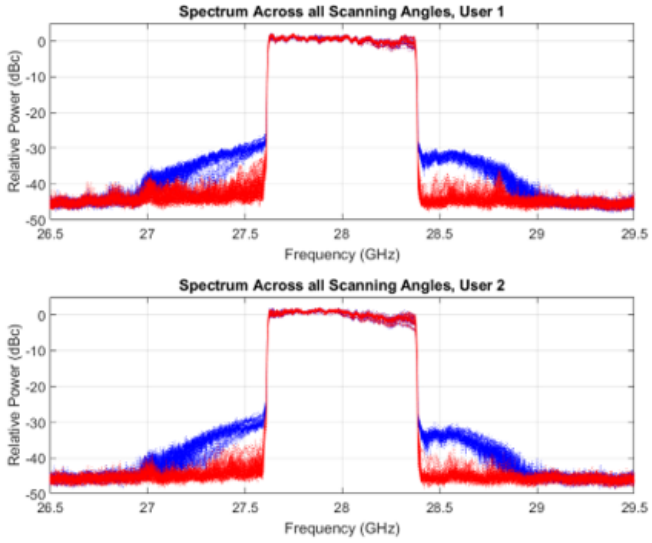

Figure 3.19: Measured spectum for user 1 (top) and user 2 (right) across all measured steering angles w/ and w/o applying set-wise $\theta_{tr} \in \{0°, 15°\}$ SISO DPDs.

Chapter 4

Summary and Conclusions

In Chapter 1, we saw how mm-wave phased arrays can be used to achieve the 100× increase in data capacity demanded by important applications such as augmented reality, driverless cars and telemedicine. We saw that the cost of deployment for these phased arrays are dependent on the EIRP they can output, and that the EIRP of current state-of-the-art phased arrays is limited by the amount of heat it generates. We then proposed the application of DPD as a method of managing the heat by allowing the array to operate further in its non-linear region where the PAs are most efficient.

In Chapter 2, we proposed a SISO model for RF beamforming arrays that was no more complex than the individual PAs in the array. The SISO model simplified the linearization task as it managed non-idealities such as coupling and load modulation as an end-to-end effect instead of a per-PA effect. From the OTA measurement results of a 64-element array, we observed that the SISO model was dependent on the beam steering angle, and we proposed a set-wise method that multiplexed between few sets of SISO coefficients to achieve good linearization across the steering range. We then developed our understanding of the origin of the non-idealities by looking at a 4-element array, in which the coupled outputs of the PAs are available. We found that uncalibrated components such as digital phase shifters can contribute to steering angle dependence. After comparing the difference between the coupled outputs from the PAs and the OTA signal, we observed that much of steering-angle dependence can be corrected using linear equalization at the receiver. Subsequently, we demonstrated that the equalizer-assisted DPD could be trained with feedback from either the coupled PA outputs or OTA outputs with similar effectiveness.

In Chapter 3, we extended the SISO model for application in multi-user RF beamforming subarrays. We proposed a training method based on linear interference estimation that

uses measurement from a single observation receiver to extract the SISO DPD coefficients for each subarray using a common feedback signal. This allowed the use of a measurement setup very similar to the single-user scenario for training the multiple SISO DPDs simultaneously without significant increase in the training signal length or computational complexity. After training the SISO DPDs with the proposed interference-canceling technique, we demonstrated the effectiveness of the SISO DPDs for improving spectral quality through comprehensive measurement results of 2 64-element subarrays across all steering angle combinations of the two users.

4.1 Future Work

The ultimate goal of this work was to demonstrate that DPD for mm-wave RF beamforming arrays is feasible under real-world measurement conditions and has the potential to be practical in commercial deployments. The presented results merely showed the feasibility of DPD in single and multi-user RF beamforming arrays, however there are many practical considerations that need to be investigated further. This includes the following avenues:

SISO DPD Coefficient Set Reduction

The complexity of the SISO DPD method can be further reduced by minimizing the required number of DPD coefficient sets and subsequently the training time. In general, this can be accomplished by making the behavior of the array more consistent across the steering angle range. Methods currently under investigation include optimization of phase shifter settings and PA bias conditions to improve the generalization of the coefficients as the beam is steered.

Open-loop vs. Closed-loop DPD

There remains ambiguity about the architecture of the TOR and even whether or not it is necessary at all. Open-loop implementations of DPD can be viewed in a similar perspective as factory calibration which would ideally only need to be trained once. It is still unclear if this can be accomplished by use of better modeling of PAs, such as by use of statistical models, or by better understanding of other aspects such as the stability of DPD coefficients over time. In the closed-loop implementation, cost-effective methods to obtain the feedback signal on arrays deployed on the field must be investigated. One

possible method being considered uses near-field probes that can be built on-board with the array front-end.

Real-time Implementation

Lastly, the training and execution of the DPD function must be implemented in real-time if it is to be deployed on production systems. There is still much study that can be done on efficient real-time implementations and characterizing the actual trade-off in power-efficiency and system up-time offered by the DPD technique.

References

[1] T. Rappaport, R. Heath, R. Daniels, and J. Murdock, *Millimeter wave wireless communications*. Prentice Hall, 2015, includes bibliographical references (pages 585-651) and index.

[2] Use of Spectrum Bands Above 24 GHz For Mobile Radio Services, et al. [Online]. Available: https://www.fcc.gov/document/spectrum-frontiers-ro-and-fnprm

[3] World Radiocommunication Conference 2015 (WRC-15), Geneva, Switzerland, 2-27 November 2015. [Online]. Available: https://www.itu.int/en/ITU-R/conferences/wrc/2015/Pages/default.aspx

[4] C. A. Balanis, *Antenna Theory: Analysis and Design*. New York, NY, USA: Wiley-Interscience, 2005.

[5] Anokiwave AWMF-0129 5G 28 GHz Active Antenna Innovator's Kit. [Online]. Available: http://www.anokiwave.com/products/awmf-0129/index.html

[6] Anokiwave AWMF-0134 5G 28 GHz 256-Element Active Antenna Innovator's Kit. [Online]. Available: http://www.anokiwave.com/products/awa-0134/index.html

[7] Cisco Visual Networking Index: Forecast and Methodology, 2016 - 2021. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf

[8] IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2083-0-201509-I!!PDF-E.pdf

[9] O. Onireti, M. A. Imran, J. Qadir, and A. Sathiaseelan, "Will 5g see its blind side? evolving 5g for universal internet access," *CoRR*, vol. abs/1603.09537, 2016. [Online]. Available: http://arxiv.org/abs/1603.09537

[10] Mobile Edge Computing in 5G Networks. [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf

[11] Towards fully connected vehicles: Edge computing for advanced automotive communications. [Online]. Available: http://5gaa.org/wp-content/uploads/2017/12/5GAA_T-170219-whitepaper-EdgeComputing_5GAA.pdf

[12] Q. Zhang, J. Liu, and G. Zhao, "Towards 5g enabled tactile robotic telesurgery," *CoRR*, vol. abs/1803.03586, 2018. [Online]. Available: http://arxiv.org/abs/1803.03586

[13] A. Al-Shuwaili and O. Simeone, "Energy-efficient resource allocation for mobile edge computing-based augmented reality applications," *IEEE Wireless Communications Letters*, vol. 6, no. 3, pp. 398–401, June 2017.

[14] A. Zanella *et al.*, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[15] Auction 73: 700 MHz Band Fact Sheet. [Online]. Available: https://www.fcc.gov/auction/73/factsheet

[16] AWMF-0129 5G 28 GHz Active Antenna Innovator's Kit. [Online]. Available: http://www.anokiwave.com/products/awmf-0129/index.html

[17] AWA-0134 5G 28 GHz 256-Element Active Antenna Innovator's Kit. [Online]. Available: http://www.anokiwave.com/products/awa-0134/index.html

[18] NR; Base Station (BS) radio transmission and reception. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3202

[19] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for rf power amplifier behavioral modeling and digital predistortion," *IEEE Trans. on Microw. Theory and Techn.*, vol. 59, no. 4, pp. 913–923, April 2011.

[20] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of rf power amplifiers," *IEEE Trans. on Microw. Theory and Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec 2006.

[21] D. R. Morgan *et al.*, "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.

[22] Signal Fitting with Subsample Resolution. [Online]. Available: https://www.dsprelated.com/showcode/207.php

[23] B. Sadhu *et al.*, "7.2 a 28ghz 32-element phased-array transceiver ic with concurrent dual polarized beams and 1.4 degree beam-steering resolution for 5g communication," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb 2017, pp. 128–129.

[24] H. T. Kim *et al.*, "A 28ghz cmos direct conversion transceiver with packaged antenna arrays for 5g cellular system," in *2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, June 2017, pp. 69–72.

[25] S. Lee *et al.*, "Digital predistortion for power amplifiers in hybrid mimo systems with antenna subarrays," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015, pp. 1–5.

[26] N. Tervo *et al.*, "Digital predistortion of amplitude varying phased array utilising over-the-air combining," in *2017 IEEE MTT-S International Microwave Symposium (IMS)*, June 2017, pp. 1165–1168.

[27] X. Liu *et al.*, "Beam-oriented digital predistortion for 5g massive mimo hybrid beamforming transmitters," *IEEE Trans. on Microw. Theory and Techn.*, vol. 66, no. 7, pp. 3419–3432, July 2018.

[28] M. Abdelaziz *et al.*, "Digital predistortion for hybrid mimo transmitters," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 445–454, June 2018.

[29] H. Yan and D. Cabric, "Digital predistortion for hybrid precoding architecture in millimeter-wave massive mimo systems," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 3479–3483.

[30] K. Hausmair *et al.*, "Prediction of nonlinear distortion in wideband active antenna arrays," *IEEE Trans. on Microw. Theory and Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov 2017.

[31] ——, "Digital predistortion for multi-antenna transmitters affected by antenna crosstalk," *IEEE Trans. on Microw. Theory and Techn.*, vol. 66, no. 3, pp. 1524–1535, March 2018.

[32] G. Z. E. Nashef *et al.*, "Em/circuit mixed simulation technique for an active antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 10, pp. 354–357, 2011.

[33] M. Romier *et al.*, "Load-pull effect on radiation characteristics of active antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 7, pp. 550–552, 2008.

[34] C. Fager *et al.*, "Analysis of nonlinear distortion in phased array transmitters," in *2017 Integrated Nonlinear Microwave and Millimetre-wave Circuits Workshop (INMMiC)*, April 2017, pp. 1–4.

[35] F. M. Barradas *et al.*, "Power, linearity, and efficiency prediction for mimo arrays with antenna coupling," *IEEE Trans. on Microw. Theory and Techn.*, vol. 65, no. 12, pp. 5284–5297, Dec 2017.

[36] N. Fatema *et al.*, "Massive mimo linear precoding: A survey," *IEEE Systems Journal*, pp. 1–12, 2017.

[37] Verspecht and D. E. Root, "Polyharmonic distortion modeling," *IEEE Microwave Magazine*, vol. 7, no. 3, pp. 44–57, June 2006.

[38] Anokiwave AWMF-0108 28 GHz Silicon 5G Tx/Rx Quad Core IC. [Online]. Available: http://www.anokiwave.com/products/awmf-0108/index.html

[39] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[40] S. Malkowsky, J. Vieira, L. Liu, P. Harris, K. Nieman, N. Kundargi, I. C. Wong, F. Tufvesson, V. Öwall, and O. Edfors, "The world's first real-time testbed for massive mimo: Design, implementation, and validation," *IEEE Access*, vol. 5, pp. 9073–9088, 2017.

[41] O. E. Ayach *et al.*, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, March 2014.

[42] S. Amin *et al.*, "Behavioral modeling and linearization of crosstalk and memory effects in rf mimo transmitters," *IEEE Trans. on Microw. Theory and Techn.*, vol. 62, no. 4, pp. 810–823, April 2014.

[43] S. Choi and E. Jeong, "Digital predistortion based on combined feedback in mimo transmitters," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1572–1575, October 2012.

[44] K. Hausmair *et al.*, "Modeling and linearization of multi-antenna transmitters using over-the-air measurements," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–4.

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

[46] T. L. Marzetta *et al.*, *Fundamentals of Massive MIMO*. Cambridge University Press, 2016.