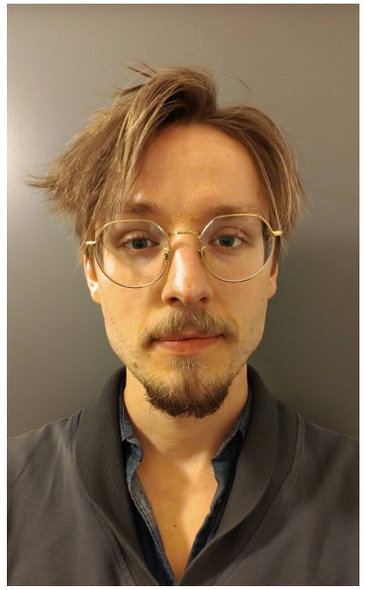
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

PHD THESIS



Communication system designs strive to constantly push the limits for higher data rates, energy efficiency, greater coverage and better quality of service, to enable new applications. Radio communication and localization in new ways, and between an ever growing number of devices, places new demands on the radio infrastructure. For radio transmitters to meet these growing demands, careful consideration of radio architectures, hardware energy consumption and imperfections, spectrum management, and much more must be done. For 5G and 6G, a key technology enabler is multi-input, multi-output (MIMO) radio designs. As for the hardware and spectrum management in the transmitters, the power amplifiers (PAs) are traditionally known as the most power consuming components. These PAs are always designed with a trade-off between power consumption and nonlinearity, from which nonlinear distortion can arise that contaminate the spectrum and deteriorates performance if not addressed. To alleviate this trade-off, digital predistortion (DPD) is a common approach to improving linearity whilst allowing the PAs to operate more efficiently.

In this thesis, DPD adaptions in various MIMO multi-user linearization scenarios are explored. With MIMO and multi-user transmitters, state-of-the-art research and this thesis argues for that traditional single-input, single-output (SISO) techniques for linearization are often superfluous and inflexible in their computational complexity, and thereby excessive in their energy consumption. With the additional degrees of freedom, and oftentimes sparsity, available in multi-user and MIMO transmitters, this thesis details existing and new techniques on how sufficient linearization can be achieved at a much lower computational cost.

CHALMERS
UNIVERSITY OF TECHNOLOGY

BJÖRN LANGBORN • Digital predistortion for power amplifier linearization in sparse multi-user systems • 2026

Digital predistortion for power amplifier linearization in sparse multi-user systems

BJÖRN LANGBORN

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Digital predistortion for power amplifier linearization in sparse multi-user systems

Björn Langborn

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden, 2026

Digital predistortion for power amplifier linearization in sparse multi-user systems

Björn Langborn
ISBN 978-91-8103-382-3

Acknowledgements, dedications, and similar personal statements in this thesis, reflect the author's own views.

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5839
ISSN 0346-718X

Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000

Cover:
Radio basestation, with seagulls perching on top, in foreground to urban landscape. Cover image credit to Frida Johansson.

Printed by Chalmers Digital Printing
Gothenburg, Sweden, March 2026

Digital predistortion for power amplifier linearization in sparse multi-user systems

Björn Langborn
Department of Electrical Engineering
Chalmers University of Technology

Abstract

To meet demands on higher data rates, ubiquitous coverage and a good quality of service in wireless communications, the trend in construction of wireless transmitters is towards multiple-input, multiple-output (MIMO) designs. In the transmitter, the power amplifier (PA) is traditionally the most power consuming component. These PAs are always designed with a trade-off between power consumption and nonlinearity, from which nonlinear distortion can arise that deteriorates performance if not addressed.

In this thesis, digital predistortion (DPD) adaptions in various multi-user linearization scenarios are explored. With MIMO and multi-user transmitters, traditional single-input, single-output (SISO) techniques for linearization are often superfluous and inflexible in their computational complexity, and thereby excessive in their energy consumption. With the additional degrees of freedom, and oftentimes sparsity, available in multi-user and MIMO transmitters, new techniques can achieve sufficient linearization at a much lower computational cost.

In Paper A, sparse concurrent multi-band transmission using a technique known as frequency relocation is studied, expanding upon previous research by considering PA gain and phase variations through linear pre-equalization filters. The combination of techniques allows for sample rate reductions in the digital-to-analog converters, resulting in reduced power consumption, whilst achieving comparable results to conventional SISO linearization.

In Paper B, concurrent multi-beam transmission in MIMO systems is explored. It is then shown that for a sparse beam space, i.e. when there are few beams to be transmitted compared to the number of transmit branches, a dimensionality reduction can be performed in a in a so-called virtual array, allowing few DPDs to linearize many transmit branches.

In Paper C, spatial sparsity is utilized for so-called spatial and frequency distortion shaping. This paper presents a technique, and a new metric, that for a fixed DPD complexity allows for improving linearization performance in selected spatial regions and for particular frequencies, at the expense of shaping nonlinear distortion to end up in directions deemed less important.

Using the techniques proposed in these papers, good communication quality can be maintained whilst operating PAs more efficiently and whilst scaling to more difficult multi-user cases.

Keywords: Digital predistortion, DPD, linearization, MIMO, sparse.

List of Publications

This thesis is based on the following publications:

[A] **Björn Langborn**, Christian Fager, Rui Hou, Thomas Eriksson, “Improved Digital Predistortion for Concurrent Multiband Transmission Using Frequency Relocation”. Published in IEEE Microwave and Wireless Technology Letters, March 2023.

[B] **Björn Langborn**, Christian Fager, Rui Hou, and Thomas Eriksson, “Concurrent multi-beam digital pre-distortion using FFT beamforming and virtual arrays”. Submitted to Sensors, special issue on "Advancements in Power Amplifier Design and Linearization Techniques for Wireless Communication Systems". Available as preprint, doi: 10.20944/preprints202602.0713.v1.

[C] **Björn Langborn**, Siqi Wang, Thomas Eriksson, “Distortion shaping in space and frequency using digital predistortion in MIMO transmitters”. Submitted to IEEE Communications Letters.

Contents

Abstract ii

List of Papers iii

Acknowledgements ix

Acronyms x

I Overview 1

1 Overview 3
1.1 Background and motivation 3
1.2 Thesis outline 6
1.3 Notation 7

2 Wireless communication system modeling 9
2.1 The transmitter 9
Modulation 10
Radio front-end architectures 13
2.2 The channel 15
Line-of-sight channel 15

Attenuation 16
Doppler effect 16
Multi-path channel 16
Channel modeling and precoding 17
Antennae 18
2.3 The receiver 19

3 Power amplifier modeling 23
3.1 Physical impairments in power amplifiers 23
3.2 Approaches to power amplifier modeling 24
3.3 Analog power amplifier models 24
The Saleh power amplifier model 25
The Volterra series 25
3.4 The baseband equivalent Volterra series 26
3.5 Power amplifier related metrics 27
Signal-to-noise and distortion ratio 27
Peak-to-average ratio 28
Adjacent channel leakage ratio 28
Normalized mean square error 30

4 Linearization using single-input, single-output digital predistortion 31
4.1 Role of digital predistortion 31
4.2 Invertability of nonlinear PAs 32
4.3 Digital predistortion models 33
Wiener and Hammerstein models 33
Look-up tables 34
Decomposed vector rotation-based DPD 34
Volterra-based DPD models 35
Frequency-domain digital predistortion 35
4.4 Digital predistortion identification 36

5 Digital predistortion in multiple-input, multiple-output settings 39
5.1 Multi-band digital predistortion 39
Multi-band transmission through nonlinear system 40
Multi-dimensional digital-predistortion methods 41
Frequency relocation 42

5.2 Multi-beam digital predistortion . . . 43
Multinomial theory to determine beam domain sparsity . . . 45
Beam-domain digital predistortion . . . 46
Virtual array digital predistortion . . . 47
5.3 Digital predistortion using over-the-air observations . . . 50
Distortion shaping in space . . . 50
Distortion shaping in frequency . . . 53
Distortion shaping in both frequency and space . . . 53

6 Summary of included papers **55**
6.1 Paper A . . . 55
6.2 Paper B . . . 56
6.3 Paper C . . . 56

7 Concluding Remarks and Future Work **59**

References **61**

II Papers 69

A Improved Digital Predistortion for Concurrent Multiband Transmission Using Frequency Relocation **A1**
1 Introduction . . . A3
2 Frequency relocation system description . . . A4
3 Proposed DPD schematic . . . A5
4 Experimental setup and results . . . A7
5 Conclusion . . . A10
References . . . A11

B Concurrent multi-beam digital pre-distortion using FFT beamforming and virtual arrays. **B1**
1 Introduction . . . B3
1.1 Previous work . . . B4
1.2 Key contributions and paper organization . . . B6
2 System modeling . . . B7
2.1 Concurrent transmission of multiple beams . . . B8
2.2 Fast-fourier transform beamforming . . . B10

3 Per-PA DPD . . . B11
3.1 Per-PA DPD example for three concurrent users . . . B11
3.2 Array aliasing . . . B13
4 Virtual array DPD . . . B13
4.1 Virtual array DPD example for three concurrent users . B14
4.2 Special beam configuration cases . . . B15
4.3 Non-DFT beamforming . . . B16
4.4 Generalization to higher order terms . . . B17
4.5 Analogy with frequency relocation . . . B18
5 Numerical setup and results . . . B18
5.1 DPD identification . . . B19
5.2 Tri-beam scenario . . . B20
5.3 Virtual array dimensioning . . . B22
5.4 Power amplifier variations . . . B24
5.5 Non-FFT precoder . . . B25
6 Conclusion . . . B26
A Finding a virtual DPD space . . . B27
References . . . B30

C Distortion shaping in space and frequency using digital predistortion in MIMO transmitters C1
1 Introduction . . . C3
2 Motivation and contribution . . . C4
3 System modeling . . . C5
4 Spatial and frequency weighted DPD . . . C7
4.1 Conventional SISO DPD . . . C7
4.2 Criteria for DPD optimization . . . C7
5 Identification of the spatially and frequency weighted DPD . . C9
5.1 Single beam . . . C10
5.2 Multiple beams . . . C10
6 Simulation setup and results . . . C11
7 Discussion . . . C14
8 Conclusion . . . C14
References . . . C15

Acknowledgments

This thesis being the product of five years of work at Chalmers, a number of people have come and gone, and been part of my existence during the process. Some helpful to the work, some helpful for not thinking of work, and some perhaps not helpful at all. In this section, I thank those from the first two aforementioned categories.

Thomas Eriksson, being my main supervisor, clearly merits first credit. Through his enthusiasm and commitment to exploring new ideas, backed by all his years of experience, he has helped me greatly in developing as a researcher. Christian Fager, my co-supervisor and previous Master thesis superisor, deserves many thanks as well, for his help in many aspects. From providing different and interesting perspectives in research meetings, to paper organization and feedback, to help with experimental matters. As for my collaborators, a first thanks goes to Rui Hou at Ericsson. Since first meeting at a digital presentation of my Master thesis work, and at all the meetings thereafter, he has always been able to provide thoughtful and intelligent research perspectives making good work great. Siqi Wang merits many thanks as well. He has helped me from first when I started as a PhD with experimental work, and has later been an appreciated co-conference attender and valuable co-author for more theoretical work. Lastly, thanks to Han Zhou, for the joint publications and time in the lab together!

As for colleagues, a first note of gratitude goes towards my office mates, Carl, José and Yu. Carl for many good work and non-work discussions in the office, and outside. José and Yu for tolerating said work discussions between me and Carl, and for many nice fika breaks! A second note goes generally towards all the colleagues at ComSys, for having provided a nice office environment, and for many good social days and alike.

As for colleagues in teaching, Erik Ström and Erik Agrell both merit special notes of thanks. Erik Ström for having been great in the role of examiner in our course on probability theory, and Erik Agrell for many patient and helpful replies when it came to my administrative duties, alongside the course on electronic systems. Also thanks to Morteza Barzegar Astanjin, for our TA work together, and pulling through when course circumstances were tough!

My thanks for the second category of people, those not from work, will have to be brief. Many thanks to family and friends, who have listened to complaints and excitement alike, during these past years!

Acronyms

ACLR:	Adjacent channel leakage ratio
ADC:	Analog-to-digital converter
ASK:	Amplitude shift keying
AWGN:	Additive white Gaussian noise
BD:	Beam-domain
CFR:	Crest factor reduction
DAC:	Digital-to-analog converter
DFT:	Discrete Fourier transform
DPD:	Digital predistortion
DVR:	Decomposed vector rotation
FD:	Frequency-domain
FFT:	Fast Fourier transform
FIR:	Finite impulse response
FSK:	Frequency shift keying
GMP:	Generalized memory polynomial
IB:	In-band
ILC:	Iterative learning control
IM:	Intermodulation
IQ:	In-phase, quadrature-phase
LO:	Local oscillator
LOS:	Line-of-sight

LUT:	Look-up table
MIMO:	Multiple-input, multiple-output
MP:	Memory polynomial
NMSE:	Normalized mean square error
OOB:	Out-of-band
OTA:	Over-the-air
PA:	Power amplifier
PAPR:	Peak-to-average power reduction
PP:	Per power amplifier
PSK:	Phase shift keying
QAM:	Quadrature amplitude modulation
RF:	Radio frequency
RIMP:	Rich isotropic multipath
Rx:	Receiver
SFW:	Spatial and frequency weighted
SISO:	Single-input, single-output
SNDR:	Signal-to-noise and distortion ratio
SNR:	Signa-to-noise ratio
Tx:	Transmitter
VA:	Virtual array

Part I

Overview

CHAPTER 1

Overview

1.1 Background and motivation

In the construction of communication systems, theoretical limits on achievable information transference between two nodes are always limited by practical impairments, physical reality and implementation costs. The impairments can range from fiber imperfections in optical networks, attenuation and multipath effects in acoustic underwater communication, Doppler shifts in satellite communication, to clock synchronization in distributed wireless systems. The examples showcase two main components limiting information transfer: The medium of propagation, as well as the implementation and equipment for transmission and reception. The effects and impairments typically affect the amplitude, phase or frequency of a signal during transmission, propagation or reception in a way that is not trivial to predict once the transmitted information is to be decoded at the receiver. Thus, various compensation methods need to be adapted. These will depend on the mode and means available for communication, in order to reach as close as possible to the theoretical limits on communication, to a reasonable cost. In this thesis, wireless radio frequency (RF) networks are of particular interest.

Several real-world examples tell the story of whether or not there truly is a need to push our practical communication systems towards their theoretical limits. The number of people staying connected with mobile devices grows linearly, and with an exponential data traffic growth [1]. Wireless applications previously unfeasible due to low data rates, physical size limitations or long communication delays, emerge as possible with the deployment of 5G and future networks. Examples include:

- Smart ambulances, where patients can receive enhanced treatment by connecting to remote medical experts and diagnostics tools, whilst in transit.

- Disaster and rescue operations using unmanned aerial vehicles, streaming video footage to site responders for quicker situation assessment.

- Inter-vehicular communication between cars on the road, or trucks in industrial environments, to decrease risk of human harm in such settings.

These applications and their wireless systems needs to be operable over a range of variable distances, with users static or moving, with potentially complex environmental conditions, and often for long durations of time. As such, demands on system design are multifaceted and stringent, resulting in that system imperfections must be treated with care. However, whilst the performance requirements are high to support these applications, they are not in effect everywhere, or at all times. For example, a radio basestation with the capacity to serve many ambulances, might at most times only have a few in its vicinity. In other words, practical applications may often operate under sparse conditions.

In wireless RF system design, the limiting imperfections for high capacity communication vary with frequency. At low frequencies, decades of research have improved equipment and electronics design to enable high power and highly efficient transmitters, as well as low-noise receivers. On the other hand, the available bandwidth is limited both physically by the carrier frequency, but also by spectrum congestion. Simultaneously, the physical channel is characterized by relatively low atmospheric attenuation and thus higher degrees of multipath components. At higher radio frequencies, hardware impairments are typically more severe, limited by available semiconductor materials and processing technologies. On the other hand, available bandwidth is more

abundant, and the wireless channel affected more by atmospheric attenuation. Notably, the dominant impairment effects, the necessary compensation techniques and available strategies for maximizing information transfer can vary with frequency.

One set of hardware impairment effects, present in all of the aforementioned examples so far to a lesser or greater extent, is power amplifier nonlinearities and memory effects. Compensation techniques for these include various schemes of feedforward, feedback or predistortion networks. The general goal of such implementations are to reduce distortion, introduced by the nonlinear and memory effects of the amplifier, below a certain threshold without e.g. excessive power consumption, or physical footprint, in relation to the application. In short, whilst feedforward networks are intuitive in theory, they typically suffer from introducing other analog imperfections in implementation, as well as introducing a large power and complexity overhead. Feedback networks on the other hand, are typically limited by stability concerns and inability to process wide bandwidths [2]. This leaves predistortion methods, implemented either in the analog or the digital domain. Such methods search for an optimal way to distort a signal prior to a nonlinear device, such that the resulting output is more linear.

Modern RF communication systems are more and more moving towards higher frequencies. This partly in order to utilize the larger bandwidths available, but the trend also comes with the benefit of making large multiple-input, multiple-output (MIMO) transceivers more practically viable from a physical dimensioning standpoint. Higher frequencies, corresponding to smaller wavelengths, lead to smaller electromagnetic components and devices. What this allows for is an added dimension to the wireless system design, namely physical space, aside from the aforementioned frequency, amplitude and phase. This because with multiple-output transceivers, the phenomena of constructive and destructive interference of waves can be utilized to spatially direct electromagnetic energy to certain directions. This is referred to as beamforming. Correspondingly, a receiver might direct its receptivity to a certain spatial region. The added dimension and ability to beamform gives flexibility, and naturally increases system sparsity in most applications. Exemplified in the earlier example with ambulances, a wireless transmitter serving few users will rarely need to utilize all its capabilities - multiplexing amplitude, frequency, phase and space to their limits - to transmit the necessary information at a rea-

sonable cost. As such, many available dimensions will oftentimes be utilized sparsely. Furthermore, for MIMO transmitters, reducing the linearization cost is not only desirable, but a necessity. With radio front-end scaling from having a few, high-power PAs to a great many medium-power PAs [3], [4], increasing linearization costs in proportion to the number of PAs becomes unfeasible. This all yields new settings for DPD research, which historically mostly has been focused on SISO systems with less inherent sparseness, namely how to utilize sparseness to lower the cost of linearization.

Contributions: This thesis contributes in the research area of digital predistortion for PAs in sparse multi-user systems. The contributions lie firstly, with Paper A, of frequency sparse, concurrent multi-band transmission to multiple users. Secondly, spatial sparsity is utilized in Paper B and Paper C, in two different ways. In Paper B, it is shown how the sparsity in the spatial domain, also referred to as the beam domain, can be used to reduce DPD complexity in larger transmit arrays. In Paper C, on the other hand, spatial sparsity is shown possible to use by shaping residual distortion to spatial and frequency regions less utilized, to lower linearization costs.

1.2 Thesis outline

This thesis is outlined as follows:

- In Chapter 2, a simplified overview of wireless communication system modeling is presented. It consists of describing the transmitter unit, the channel present in wireless radio, and the receiver. Hardware impairments are commented upon, as they are of particular interest in this thesis.

- In Chapter 3, power amplifier modeling is detailed. Specifically, how impairments can be modeled in both an analog continuous time, and a baseband digital, description is elaborated upon. Metrics of relevance for stating the degree of impairments presents are provided.

- In Chapter 4, digital predistortion in the context of single-input, single-output systems is addressed. The purpose, theory and common models for digital predistortion are provided, alongside common methods of identification.

- In Chapter 5, digital predistortion is instead considered in the context of multi-input, multi-output systems. Specifically, multi-band and multi-beam signals are considered. State-of-the-art techniques and researched is given in context for the novel ideas presented in this thesis.
- Finally, Chapter 6 summarizes the papers that make up the foundation of this thesis, and Chapter 7 provides some concluding remarks and comments on future work.

1.3 Notation

The mathematical notation in this work is as follows. Bold lower case symbols, e.g. $\boldsymbol{h}$, $\boldsymbol{x}$, denote vectors, whilst bold upper case symbols such as $\boldsymbol{H}$ denote matrices. Subscripts, e.g. $\boldsymbol{x}_l$, serve as indices for e.g. multiple signals $\boldsymbol{x}_1, ..., \boldsymbol{x}_L$ on different branches of an array. Superscripts either denote indexing in space, if contained in parenthesis such as $\boldsymbol{y}^{(k)}$, or iteration indices if contained within square brackets such as $\boldsymbol{x}^{[i]}$.

CHAPTER 2

Wireless communication system modeling

Wireless communication systems are typically modeled in blocks. At a high level, they are often decomposed into the transmitter unit, the receiver and the channel between. However, within these blocks, several different architectures or models can be assumed or implemented, with a variety of components or underlying theory. In this section, a brief overview is presented, followed by an example of how the transmitter and receiver operates schematically. Particular emphasis on transmitter impairments will be placed.

2.1 The transmitter

A wireless transmitter can be implemented in various ways, but most implementations share the following fundamental building blocks in order:

1. An information source, digital or analog data producing baseband data that an application wishes to transmit.

2. Source and channel coding, for compressing the baseband data and making it robust towards errors induced by the channel.

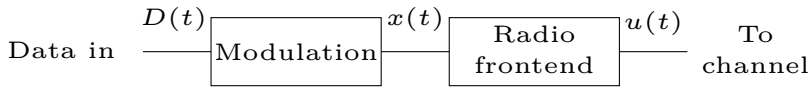


Figure 2.1: Block schematic of transmitter, from input data bits, through modulation and radio frontend.

3. Modulation, where more efficient use of the radio spectrum can be achieved by modulating a baseband signal onto a carrier frequency or to a different format.

4. A radio frontend, where mixing, filtering, frequency conversion and amplification of an analog signal is performed.

For the purposes of this work, little needs to be known of the exact details of how the source information is coded into digital bits. Thus, a relevant truncation of scope is to assume a series of digital bits is given to a modulation, as in Fig. 2.1. However, the remaining building blocks need further detailing.

Modulation

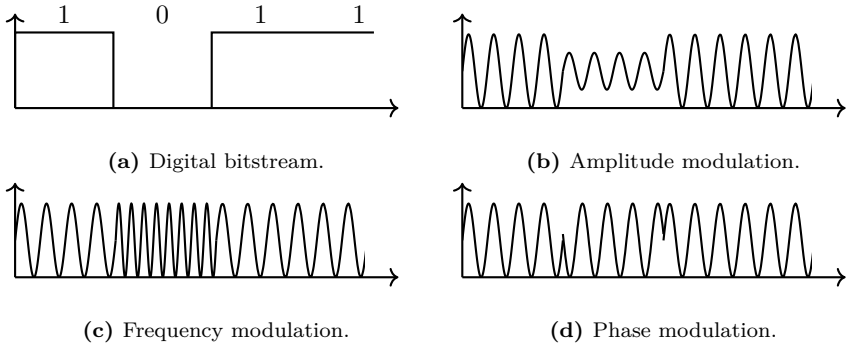
Signal modulation can be done either mostly in the analog domain, or mostly in the digital domain. However, whilst analog modulation schemes have been applied widely historically, and still sees use in some applications today, digital modulation schemes are now far more common. This due to better spectral efficiency, better channel performance and compatibility with coding schemes [5]. In essence though, regardless of analog or digital implementation, the modulation principles are similar. Consider Fig. 2.1, and let the signal as a result of modulation be represented as time-continuous signal that at time t is

$$x(t) = f_{\mathrm{MOD}}(D(t))\,. \tag{2.1}$$

If the input data is an analog baseband waveform, $f_{\mathrm{MOD}}(\cdot)$ can incorporate different functions, e.g. frequency modulation, as

$$f_{\mathrm{MOD}}(D(t)) = A\cos\left(2\pi f_{\mathrm{I}} t + k\int_{\tau=0}^{t} D(\tau)d\tau\right), \tag{2.2}$$

where f_{I} denotes an intermediate frequency, A an amplitude scaling and k a factor determining the rate of change in frequency. If instead $D(t)$ is a



(a) Digital bitstream. **(b)** Amplitude modulation.

(c) Frequency modulation. **(d)** Phase modulation.

Figure 2.2: Digital bitstream, and corresponding amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK) modulation schemes.

digital waveform, Fig. 2.2 visualizes three fundamental ways to modulate the signal through other functions $f_{\text{MOD}}(\cdot)$ to obtain $x(t)$, with $D(t)$ encoded into the amplitude, frequency or phase. These are simple, but inefficient in using the radio spectrum. Thus, there is a vast amount of literature on how to modulate the digital bitstreams in a more efficient manner. A first, common, method is to use the quadrature components of a signal at a given frequency. Mathematically, in baseband the modulation is

$$x(t) = x_{\text{I}}(t) + jx_{\text{Q}}(t)\,, \tag{2.3}$$

where data is encoded into the in-phase component x_{I}, and into the quadrature component x_{Q}. From this notation, the modulation is often denoted as IQ modulation.

In addition to letting data be encoded on different components, grouping bits in the digital bit stream to symbols can also be done. This leads to that each modulation format is no longer binary but rather has multiple levels. Exemplification of a scheme utilizing both these approaches is 16-QAM, visualized in the form of a constellation diagram in Fig. 2.4a. Here, groups of four bits are mapped to complex numbers, where e.g. in Fig. 2.4a the bit sequence 0000 corresponds to the complex number $-3 + j3$. Using two components, and grouping bits into symbols, will evidently increase the number of bits possible to transmit in a limited time, though at the cost of increased sensitivity to noise, distortion and other impairments when decoded at the

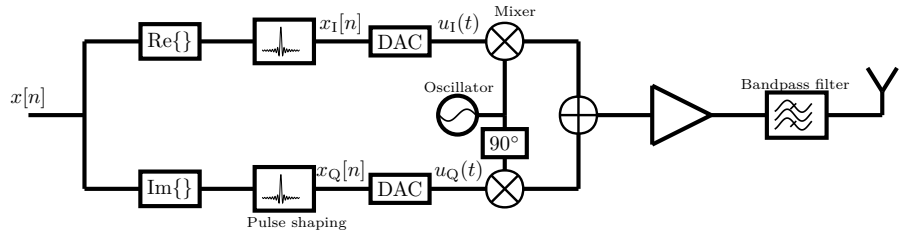
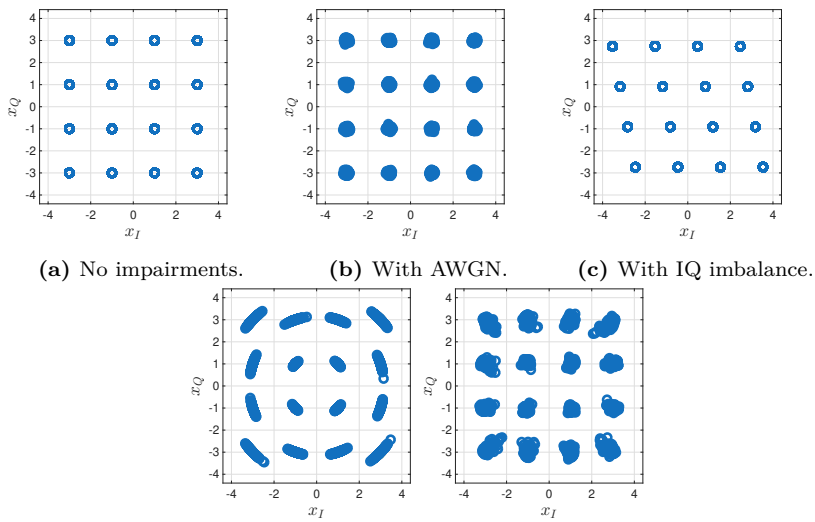


Figure 2.3: Simple front-end transmitter architecture for IQ modulated baseband data.



(a) No impairments. **(b)** With AWGN. **(c)** With IQ imbalance.

(d) With phase noise. **(e)** With nonlinearity.

Figure 2.4: Single carrier QAM constellations with various hardware impairments.

receiver. Thus, higher order modulation schemes are limited by the amount of noise, distortion and impairments present in the system. The impact of various hardware impairments will be detailed in the following subsection.

Radio front-end architectures

As indicated in the previous section, using quadrature components in modulation is one effective way to increase the data rate. In Fig. 2.3, a simple schematic is presented for how a radio frontend using IQ modulated data, i.e. with $x(t)$ as in (2.3), can be implemented. Here, x is assumed be a time discrete, sampled signal, thus denoted

$$x[n] = x_{\mathrm{I}}[n] + jx_{\mathrm{Q}}[n]\,, \tag{2.4}$$

with real and complex parts x_{I} and x_{Q}, respectively, and where n denotes a sample. The schematic otherwise consists of digital-to-analog converters (DACs), a local oscillator, 90° delay, two mixers, an adder, a power amplifier, an antenna and a bandpass filter. The sinc-like box represents upsampling and pulse shaping, whilst Re{} and Im{} represents taking the real and imaginary value of the complex signal, respectively. In what follows, the mathematical steps that detail this transmitter architecture will be explained, along with short comments on the involved components.

In the first step, the IQ modulated signal is split into its real part and imaginary part. Assuming this is a QAM modulated signal, the constellation diagram will ideally look as in Fig. 2.4a. Afterwards, these are upsampled, pulse shaped and converted from digital signals to analog, using the DACs. The signal after the DAC are then two analog signals

$$\begin{aligned} u_{\mathrm{I}}(t) &= \sum_{n=0}^{N} x_{\mathrm{I}}[n] r(t-n) \\ u_{\mathrm{Q}}(t) &= \sum_{n=0}^{N} x_{\mathrm{Q}}[n] r(t-n)\,. \end{aligned} \tag{2.5}$$

Here, $r(t)$ represents pulse shaping, to shape the spectral properties of the signal. No limitation on the number of amplitude levels of $x_{\mathrm{I}}[n]$ or $x_{\mathrm{Q}}[n]$ has been presumed here, whilst in practice DACs often limit these to 2^B discrete levels, where B typically range from the cheapest $B = 1$ implementations to about $B = 16$ for more complicated architectures. In addition, hardware imperfections can cause effects such as amplitude and phase imbalance between the generated $u_{\mathrm{I}}(t)$ and $u_{\mathrm{Q}}(t)$ components. For the purposes of this work, (2.5) suffices as a description, though Fig. 2.4c illustrates how IQ imbalance can affect the reception of the IQ data.

Mixing the analog signal components $u_{\mathrm{I}}(t)$ and $u_{\mathrm{Q}}(t)$ to a carrier frequency through the mixer produces

$$u_{\mathrm{RF}}(t) = u_{\mathrm{I}}(t)\cos\left(2\pi f_c t\right) + u_{\mathrm{Q}}(t)\sin\left(2\pi f_c t\right). \tag{2.6}$$

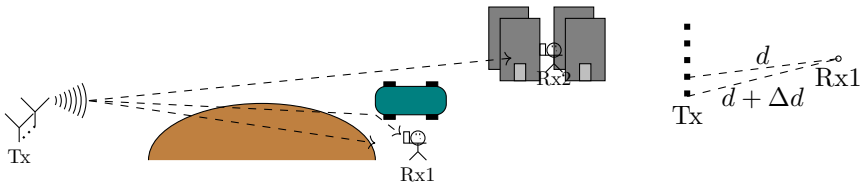
Again, this is a simplified description, as the mixers are nonlinear and non-trivial components to design and characterize. One common hardware imperfection, referred to as phase noise, is in this step to varying degrees induced from the local oscillator via the mixer into the signal. With this effect present, (2.6) would be modified to

$$u_{\mathrm{RF}}(t) = u_{\mathrm{I}}(t)\cos\left(2\pi f_c t + \phi(t)\right) + u_{\mathrm{Q}}(t)\sin\left(2\pi f_c t + \phi(t)\right), \tag{2.7}$$

where $\phi(t)$ is the phase noise, varying in time. With present and unknown phase noise, the signal might be erroneously decoded at the receiver, indicated in Fig. 2.4d.

Passing the signal of (2.6) through a nonlinear amplifier with memory effects, a nonlinearly filtered output is produced. As the PA modeling is closely linked to DPD and linearization, this will be given more careful details in Chapter 3. For the purposes of this brief overview, the function f_{PA} will simply denote the PA model, to occur later when analyzing the received signal. Depending on type of nonlinearity, signal samples can either be compressed or expanded. The effect of compressive PA behaviour on a QAM constellation is seen in Fig. 2.4e. To remove harmonic components, induced by the PA nonlinearity, from radiating, a bandpass filter is placed after the PA.

Finally, an antenna radiates the signal out from the PA over-the-air (OTA). Depending on the where the receiver is located, the signal transfer will vary. In particular, dependent on the distance, the receiver might be in the reactive or radiative near-field, or in the far-field. For the purposes of this thesis, the far-field is of foremost interest, defined when the receiver is more than the so-called Fraunhofer distance away, often computed by $2D^2/\lambda$ [6]. Here, D is the diameter of the antenna and λ the wavelength. Depending on antenna design, the characteristics of the radiated signal in the far-field, in terms of gain and phase, will vary in different directions. Thus, the antenna can be considered a linear, spatial filter prior to the channel.



(a) Environment overview.

(b) LOS path from Tx to Rx1, top view.

Figure 2.5: Channel environment with one transmit unit, Tx, and two receiving users, Rx1 and Rx2. Rx1 receives one LOS path attenuated by a hill, and one reflection path from a nearby car. Meanwhile, the transmitted signal towards Rx2 is expected to scatter multiple times within the dense housing environment before reaching the user.

2.2 The channel

Between the transmitter and receiver in a wireless communication link, there will be a channel, illustrated in Fig. 2.5. Whilst the medium, obstacles and dynamics of the environment can vary in complex ways, several relatively simple models are typically adapted to account for such variations. Which model to adapt depends on the channel and present propagation phenomena that might occur. Here, some common channel phenomena and their effect on the channel will be presented.

Line-of-sight channel

The line-of-sight (LOS) channel is the simplest one, where there is a direct path between the transmitter and receiver. Here, the attenuation and delay in signal power can computed from Friis transmission equation

$$P_{\text{rx}} = P_{\text{tx}} G_{\text{rx}} G_{\text{tx}} \left(\frac{\lambda}{4\pi R} \right)^2 , \tag{2.8}$$

where P_{rx}, P_{tx} are the transmitted and received power respectively, G_{rx}, G_{tx} the antenna gains, R the distance and λ carrier wavelength. It is from this evidently quite easy to compute the necessary power from the transmit side to reach a particular power level at the receiver. Additionally, seen from Fig.

2.5b, the relative delay from different transmit branches to a receiver location is geometrically computable. This is a key observation enabling beamforming, i.e. that the phase on the different transmit branches can be changed to steer where the signal power ends up adding constructively in space. This is revisited in a later subsection.

Attenuation

Signal attenuation occurs when the signal propagates from the transmitter through any medium, to the receiver simply due to that the power is spread across space. Higher attenuation will be incurred if the propagation goes through a medium with a higher propagation loss than air, before reaching the receiver. Such a medium can e.g. be a hill, or the constituents of a house in which a receiver is located. In addition, if the receiver or environment is changing, this attenuation can vary over time. This is then referred to as either slow, or fast, fading, dependent on the rate of change.

Doppler effect

If either the receiver, or transmitter, moves quickly, the so-called Doppler effect can occur. This effect is commonly understood from the everyday occurrence of how the sound, the frequency, from a passing ambulance changes as it drives by. Similarly at RF, the perceived effect at the receive side is that the carrier frequency varies. This change can by approximated by the formula

$$f_{\mathrm{rx}} = \frac{v_m + v_{\mathrm{rx}}}{v_m - v_{\mathrm{tx}}} f_{\mathrm{tx}} \,, \tag{2.9}$$

where f_{rx} is the received frequency, f_{tx} the transmitted frequency, v_m is the speed of the electromagnetic wave in the medium, and v_{rx}, v_{tx} the velocity component with which the receiver and transmitter moves relatively towards one another, respectively.

Multi-path channel

In practical environments a signal will typically have multiple propagation paths that can lead from the transmitter to the receiver. This is then referred to as a multi-path channel. However, the exact conditions under which this

occurs can vary drastically. For example, in a rural setting with few infrastructural elements, fewer paths are typically present. In contrast, a dense urban environment might deprive LOS conditions, but instead present many paths. One type of rich scattering channels is the so-called rich isotropic multi-path channel (RIMP), in which signals impinge upon the receiver from different directions in space isotropically.

Three main components and effects in a multi-path channel are diffractive, reflective and scattering multi-path components. These have different physical explanations for their occurrences, but for the purposes of this work, it suffices to state two things. Firstly, that these phenomena give different multi-path components with different kind of attenuative and phase properties. Secondly, that their presence in a channel will vary with frequency and with the sizes of objects in the environment.

Channel modeling and precoding

There are many channels models to account for the various channel phenomena. Oftentimes, the mathematical model is phrased as a linear filter. In a digital representation, for a single transmitter and receiver, the received signal y is then related to the transmitted signal x as

$$y[n] = \sum_{m=0}^{M} h_m x[n-m]\,. \tag{2.10}$$

If the channel has only a LOS path, no memory delays need to accounted for, yielding $M = 0$, and $|h_0|$ will follow from Friis transmission equation. A RIMP channel, in contrast, is characterized by that each h_m follows a complex Gaussian distribution, $h_m \sim \mathcal{CN}(a, \sigma)$. The parameters a and σ determine the channel gain.

The channel models are useful to obtain channel estimates. Given a channel estimate, the role of a precoder in the communication system is to utilize channel knowledge to shape how the signal is received in certain spatial positions. Considering a LOS channel with multiple transmitters, beamforming is a special kind of precoding. Let $y^{(k)}$ denote the received signal in direction (k), and let transmit branch l transmit u_l, such that

$$y^{(k)} = h_1^{(k)} u_1 + ... + h_L^{(k)} u_L\,. \tag{2.11}$$

The complex constants $h_l^{(k)}$ are determined by the geometry of the environment, where the magnitude corresponds to signal attenuation from propagation and phase corresponds to relative propagation delay. Assuming each branch transmit the same user stream x, such that $u_1 = ... = u_L = x$, then maximal constructive interference occurs when $h_1^{(k)} = ... = h_L^{(k)}$, and similarly maximal destructive interference when $h_1^{(k)} + ... + h_L^{(k)} = 0$. However, in order to shape in which directions this occurs, let $u_l = w_l x$, where w_l are beamforming weights. The choice of w_l then determines where interference occurs. Beamforming can naturally be generalized to include multiple user streams $x_1, ..., x_U$, as

$$y^{(k)} = h_1^{(k)}(w_{1,1}x_1 + ... + w_{1,U}x_U) + ... + h_L^{(k)}(w_{L,1}x_1 + ... + w_{L,U}x_U)\,, \quad (2.12)$$

where

$$u_l = w_{l,1}x_1 + ... + w_{l,U}x_U\,. \quad (2.13)$$

One computationally efficient implementation of beamforming uses the discrete Fourier transform (DFT) beamforming matrix, where given

$$\omega = e^{-j2\pi/L}\,. \quad (2.14)$$

the beamformer is formed as

$$\boldsymbol{W}_{\text{DFT}} = \frac{1}{\sqrt{L}}\begin{bmatrix} 1 & 1 & 1 & ... \\ 1 & \omega & \omega^2 & ... \\ 1 & \omega^2 & \omega^4 & ... \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = \frac{1}{\sqrt{L}}\left[\,\boldsymbol{\omega}^{(b_1)}\ \boldsymbol{\omega}^{(b_2)}\ \boldsymbol{\omega}^{(b_3)}\ ...\right]\,. \quad (2.15)$$

Here, $b_i \in [1, L]$ are beam indices, indexing the beams on the DFT-beamgrid. In implementation, the computations can be done efficiently using Fast Fourier transforms (FFTs).

Antennae

Antennae are passive RF components, meant to radiate power, converting from electrical currents to electromagnetic waves. This will typically involve designing a metallic structure dimensioned after the wavelength it is intended to radiate. It will as such be operable, have a pass-band, over a limited range of frequencies whilst acting as a stop-band filter for other frequencies. In addition, it is often designed to radiate particularly well in certain spatial

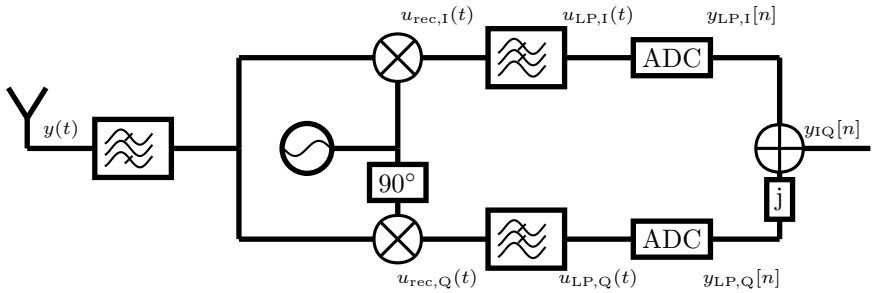


Figure 2.6: Simple front-end receiver architecture for IQ modulated baseband data.

directions, to concentrate the electromagnetic energy. As such, it acts as a filter both in frequency and space. For the purposes of this work, a model

$$a(\phi, f) = a_{\phi,f}\,, \tag{2.16}$$

suffices to describe this behavior, where ϕ is an angular direction in which an antenna radiates, f is the frequency, and $a_{\phi,f}$ a complex coefficient.

2.3 The receiver

A simple example of a receiver architecture is depicted in Fig. 2.6, consisting of a receive antenna, a bandpass filter, a local oscillator, mixers, a 90° phase shifter, lowpass filters, analog to digital converters (ADCs), and digital conversion from real to complex numbers. As part of the digital processing, another common block is equalization, to account for the different phenomena described under Section 2.2. In this section, the blocks will be detailed mathematically as done for the transmitter in Section 2.1, though fewer comments on hardware impairments.

Starting at the receive antenna, let the received signal be denoted as

$$y(t) = \int_{\tau} h(\tau) h_{\mathrm{ant,rx}} h_{\mathrm{ant,tx}} f_{\mathrm{PA}}(u_{\mathrm{RF}}(t-\tau))\mathrm{d}\tau\,, \tag{2.17}$$

where h is filter describing the channel response, $h_{\mathrm{ant,rx}}$, $h_{\mathrm{ant,tx}}$ are single scalar coefficients describing the antenna gain of the receiver and transmitter, respectively, whilst f_{PA} is the transmitter-side PA model, and u_{RF} is as in

(2.6). Typically, one also needs to include thermal noise here, as it can affect the received signal considerably. In a QAM constellation, the noise will spread the received symbols in the constellation diagram, as seen in Fig. 2.4b. For the mathematical analysis here though, thermal noise is omitted. Furthermore, let for simplicity $h(\tau) = \delta(\tau)$, which corresponds to that the channel has no memory behavior and unity gain, and let $h_{\text{ant,rx}} = h_{\text{ant,tx}} = 1$. Then (2.17) becomes

$$y = f_{\text{PA}}(u_{\text{RF}}(t)) \,, \tag{2.18}$$

i.e. the received signal is exactly what we transmitted. Applying a bandpass filter f_{BP} followed by mixing with $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ yields

$$\begin{aligned} u_{\text{rec, I}}(t) &= f_{\text{BP}}(f_{\text{PA}}(u_{\text{RF}}(t))) \cos(2\pi f_c t) \,, \\ u_{\text{rec, Q}}(t) &= f_{\text{BP}}(f_{\text{PA}}(u_{\text{RF}}(t))) \sin(2\pi f_c t) \,. \end{aligned} \tag{2.19}$$

It is realised from (2.6) that both factors $u_{\text{RF}}(t)$ and $\cos(2\pi f_c t)$, or $\sin(2\pi f_c t)$, in (2.19) will have a main component around the carrier frequency f_c. The role of the bandpass filter is then to ensure no other signals that reach this receiver are propagated in the receiver chain. From this architecture, it can be assumed that far-away carriers are filtered away, whilst all signal components sufficiently close to f_c are propagated. Let, for simplicity, the PAs operate as ideal linear amplifiers with gain g. If no other signals are received close to the carrier f_c, then (2.19) can be simplified to

$$\begin{aligned} u_{\text{rec, I}}(t) &= g u_{\text{RF}}(t) \cos(2\pi f_c t) \,, \\ u_{\text{rec, Q}}(t) &= g u_{\text{RF}}(t) \sin(2\pi f_c t) \,. \end{aligned} \tag{2.20}$$

This is an ideal description, assuming that the receiver LO frequency perfectly matches that of the transmitter. If this is not the case, i.e. if the LO instead generates $f_c + \Delta f$ at the receiver, so-called carrier frequency offset will occur.

Following the mixing, lowpass filters remove images of the signal that are induced at higher frequencies. This can readily be seen by expanding (2.20) with (2.6), such that

$$\begin{aligned} u_{\text{rec, I}}(t) &= g(u_{\text{I}}(t) \cos(2\pi f_c t + \phi(t)) + u_{\text{Q}}(t) \sin(2\pi f_c t + \phi(t))) \cos(2\pi f_c t) \,, \\ u_{\text{rec, Q}}(t) &= g(u_{\text{I}}(t) \cos(2\pi f_c t + \phi(t)) + u_{\text{Q}}(t) \sin(2\pi f_c t + \phi(t))) \sin(2\pi f_c t) \,. \end{aligned} \tag{2.21}$$

and applying the identities $\cos^2 a = 1/2 + 1/2 \cos 2a$, $\sin^2 a = 1/2 - 1/2 \cos 2a$ and $\cos a \sin a = 1/2 \sin 2a$. Remaining will just be the streams

$$\begin{aligned} u_{\mathrm{LP,I}}(t) &= \frac{1}{2} g u_{\mathrm{I}}(t) \\ u_{\mathrm{LP,Q}}(t) &= \frac{1}{2} g u_{\mathrm{Q}}(t) \,. \end{aligned} \tag{2.22}$$

In this, a narrowband assumption has been made, namely that the bandwidths of u_{I} and u_{Q} are much less than the carrier frequency f_c.

Converting from analog to digital, matched filters to the filters $r(t-n)$ in (2.5) and scaling are applied, yielding

$$\begin{aligned} y_{\mathrm{LP,I}}[n] &= 2\frac{g}{2} x_{\mathrm{I}}[n] \,, \\ y_{\mathrm{LP,Q}}[n] &= 2\frac{g}{2} x_{\mathrm{Q}}[n] \,, \end{aligned} \tag{2.23}$$

followed by combining into a single complex number

$$y_{\mathrm{IQ}}[n] = g(x_{\mathrm{I}}[n] + j x_{\mathrm{Q}}[n]) \,. \tag{2.24}$$

With this showcase of a simple transmit, channel and receive chain, the following section will delve deeper into the effect that imperfect, nonlinear power amplifiers have on the system.

CHAPTER 3

Power amplifier modeling

Power amplifier knowledge and accurate modeling is vital for digital predistortion. If an ampifier can be modeled accurately with an invertible function, the digital predistortion will seek to identify that inverse, discussed further in Chapter 4. However, for the problem to be computationally feasible in real-time application, a compromise between accuracy and complexity must be made.

In this chapter, a cursory overview of physical effects present in power amplifiers will be detailed. Following this, models capturing these effects will be explored, ending up in deriving a common model for both PA modeling and digital predistortion, the baseband equivalent Volterra series. Lastly, a few common metrics used for communication with nonlinear amplifiers are listed.

3.1 Physical impairments in power amplifiers

Most power amplifier circuits used for RF wireless communication use semiconductor devices and transistors, and as such any nonideality of the used semiconductor materials will be present in the electronic circuit. Material

properties, such as electron mobility, breakdown electric field and thermal conductivity will determine how large currents and high voltages a device can handle, and how well it can dissipate heat. Dielectric properties will affect parasitic effects [7]. Furthermore, impairments from other connected circuitry, e.g. the bias network, alongside the device structure itself and class of amplifier will also affect the output power, linearity, and the memory effects of a power amplifier circuit, and the frequency range over which the circuit can operate [8]. Lastly, varying the input and output impedances of a PA will change its behavior, an effect that is especially prominent in multi-antenna transmitter systems [9], [10].

3.2 Approaches to power amplifier modeling

A complex device such as a power amplifier can be modeled in many different ways. A first approach could be to include all known physical phenomena present in the device, to go from a given input to predict the output. These kinds of models are referred to as physically based models. A second alternative, a purely data driven approach, could be to measure the input and output in various settings, and save the result in a table as reference for future predictions. A third option, which is typically adopted for power amplifier modeling in the context of digital predistortion, is so-called behavioral modeling. Here, the general behavior going from an input the output is modeled by some governing equation. This type of model is typically based on the governing physics, but greatly simplified compared to a full description. At the same time, it incorporates data from input-output measurements to determine parameters of the model.

3.3 Analog power amplifier models

Whilst many behavioural models, such as Bessel-Fourier, neural networks or Wiener models have been used to describe PA behaviour [11]–[14], only two models used in this thesis will be detailed here. First, the Saleh model [15], and secondly the Volterra series.

The Saleh power amplifier model

The Saleh model for PA modeling is a relatively simple formula relating an RF input

$$x(t) = r(t)\cos\left(2\pi f t + \Psi(t)\right), \tag{3.1}$$

to the RF output $y(t)$ of the PA

$$y(t) = A(r(t))\cos\left(2\pi f t + \Psi(t) + \Phi(r(t))\right). \tag{3.2}$$

Here, two functions $A(\cdot)$, $\Phi(\cdot)$ characterize the transfer function at any given time t, in [15] proposed to be modelled as

$$\begin{aligned} A(r(t)) &= \frac{\alpha_a r(t)}{1+\beta_a r(t)^2}, \\ \Phi(r(t)) &= \frac{\alpha_\phi r(t)^2}{1+\beta_\phi r(t)^2}. \end{aligned} \tag{3.3}$$

Here, α_a controls the linear PA gain, β_a the static nonlinearity, whilst α_ϕ and β_ϕ will change the phase response. This model typically sees use for PA modeling with static or narrowband nonlinearities. However, it is limited in its ability to model a general PA, even in a limited operational range. In contrast, the next subsection will detail the Volterra series, that is more generally applicable.

The Volterra series

The Volterra series is a power series with memory terms, general enough in its mathematical formulation to be used in many fields of science. Generally stated, a full Volterra series, for a continuous input signal $x(t)$ to an output $y(t)$, considering memory effects within a time span T from the current time t, can be phrased as

$$y(t) = a_0 + \sum_{p=1}^{P} \int_{\tau_p=-T}^{T} \cdots \int_{\tau_1=-T}^{T} a_p(\tau_1,\ldots,\tau_p) \prod_{i=1}^{p} x(t-\tau_i)\, d\tau_i, \tag{3.4}$$

where P is the maximal nonlinearity order considered. Given this statement, a few simplifications and notes can be made with the physical knowledge of the system that is modeled. Firstly, given causality, then the integration range changes from $(-T, T)$ to $(0, T)$. Secondly, when characterizing the kernel, T

can be truncated to a finite value based on physical knowledge or measurements of short term and long term memory effects. Thirdly, it is known that the power amplifier behavior will vary, for example, with frequency, biasing and temperature. As such, each model instance, with a set kernel, as in (3.4) is only expected to be valid in the proximity of the circumstances under which they were identified. That is, the kernel a_p might need to be re-identified if the frequency, biasing or temperature changes.

Assume that a much reduced version of (3.4) suffices for modeling of a certain PA, as

$$y(t) = f_{\text{PA}}(x) = \int_\tau a_1(\tau)x(t-\tau)\mathrm{d}\tau + \int_\tau a_3(\tau)x^3(t-\tau)\mathrm{d}\tau\,. \tag{3.5}$$

The first term is a linear filtering of the input, whilst the second term adds a third order nonlinearity. If $a_1(\tau), a_3(\tau)$ are not constant, then the PA response varies with time, and memory effects occur. Here, $a_1(\tau), a_3(\tau)$ are real-valued. If $a_3(\tau) > 0$ at any given time τ then the distortion will be expansive at that time, i.e. stretch out the signal compared to if $a_3(\tau) = 0$. In contrast, and most often present in practice, $a_3(\tau) < 0$ leading to compressive behavior.

In order to characterize the model in (3.4), a corresponding digital representation will need to be stated. Furthermore, converting from analog to digital at RF is inconvenient, as it often requires a prohibitively high sampling rate. Instead, the following subsection details how the model can be formulated at baseband.

3.4 The baseband equivalent Volterra series

Different formulations of a Volterra series, pruned and truncated for practical use with baseband signals has been proposed [16]–[18]. The general approach of formulating the discrete baseband model from the continuous RF model is similar to how, in Chapter 2, equation (2.4) relates to (2.24). That is, the PA output of (3.4) is typically assumed filtered, before it is downconverted and digitized. The continous time kernel $a_p(\tau_1, \ldots, \tau_p)$ for each nonlinearity order p is approximated with a digital kernel $c_p[m_1, ..., m_p]$. One formulation then

is to go from (3.4) to

$$\begin{aligned} y[n] =& a_0 + \\ &\sum_{p=1,3,\ldots}^{P} \sum_{m_1=0}^{M} \cdots \sum_{m_p=0}^{M} c_p[m_1, ..., m_p] x[n-m_1] \cdot \\ &\prod_{i_1=0}^{(p-1)/2} x[n-m_{1+i_1}] \prod_{i_2=0}^{(p-1)/2} x^*[n-m_{(p-1)/2+1+i_2}] \,. \end{aligned} \tag{3.6}$$

Notable in (3.6) is the inclusion of complex conjugate terms x^*. These arise from the derivations since in (3.6), x is a complex-valued baseband signal, whereas in (3.4), x is a real-valued RF signal.

3.5 Power amplifier related metrics

In this subsection, a few common metrics for evaluating the communication system link quality, when power amplifiers are present, are listed. These relate the desired signal and system behavior to, in particular, nonlinear hardware impairments. Furthermore, with (3.6) at hand, the thesis will from this point on focus on digital baseband signal representations. Thus, the metrics will be listed accordingly.

Signal-to-noise and distortion ratio

Let $\boldsymbol{x} = \begin{bmatrix} x[1] & \ldots & x[N] \end{bmatrix}^T$ be the signal of interest for a transmitter. Assume then that $\boldsymbol{x}$ is passed through a nonlinear amplifier $f_{\mathrm{PA}}(\cdot)$, and is subject to thermal noise $\boldsymbol{\sigma}$, as

$$\boldsymbol{y} = f_{\mathrm{PA}}(\boldsymbol{x}) + \boldsymbol{\sigma} \,. \tag{3.7}$$

By a Bussgang decomposition of $\boldsymbol{y}$, as in [19], (3.7) can be represented as

$$\boldsymbol{y} = g\boldsymbol{x} + \boldsymbol{d} + \boldsymbol{\sigma} \,, \tag{3.8}$$

where g is the useful gain of the amplifier, and $\boldsymbol{d}$ is distortion uncorrelated to $\boldsymbol{x}$. From (3.8), the signal-to-noise and distortion ratio at the PA output can be stated as

$$\mathrm{SNDR} = \frac{g^2 \boldsymbol{x}^H \boldsymbol{x}}{\boldsymbol{\sigma}^H \boldsymbol{\sigma} + \boldsymbol{d}^H \boldsymbol{d}} \,. \tag{3.9}$$

Setting $\boldsymbol{d} = \mathbf{0}$, (3.9) yields the signal-to-noise ratio (SNR). Furthermore, considering the residual distortion power as noise, this metric can be related, via the Shannon-Hartley theorem, to the maximum capacity of the system.

Peak-to-average ratio

The variability of a signal $\boldsymbol{x}$, how high the peak amplitudes are in relation to the typical signal magnitudes, can be quantified and expressed in terms of the peak-to-average power ratio

$$\text{PAPR} = \frac{\max\limits_{1 \leq n \leq N} |x[n]|^2}{\sum_{n=1}^{N} |x[n]|^2} \,. \tag{3.10}$$

This metric is important in determining how nonlinearities will appear in the system. Typically, if the PAPR is high, the power amplifier can not be driven with too high mean input power. The practice then of lowering the mean input power, to account for high PAPR, is referred to as going into back-off, in order to avoid strong nonlinear effects.

A practical concern with (3.10) lies in the potential extreme values of $|\boldsymbol{x}|^2$. Consider that $x[n] \sim \mathcal{CN}(0, 1)$. For an infinite number of samples, $|x[n]|^2$ can be arbitrarily large, thus making the metric useless. For a finite number of samples, on the other hand, it is possible that a single sample is drawn with extremely high magnitude, whilst the remainder are closer to the signal mean. To account for these concerns, a practical approach can be to not choose the single largest sample magnitude, but instead sort samples with respect to magnitude and choose the one representing e.g. the 99.9:th% quantile. Mathematically,

$$\begin{aligned} \text{PAPR} &= \frac{|x[\tilde{n}]|^2}{\sum_{n=1}^{N} |x[n]|^2} \,, \\ \tilde{n} &= \arg \min_{1 \leq n \leq N} |F_{|\boldsymbol{x}|}(x[\tilde{n}]) - 0.999| \,, \end{aligned} \tag{3.11}$$

where $F_{|\boldsymbol{x}|}(x[n])$ is the empirical distribution function of $|\boldsymbol{x}|$.

Adjacent channel leakage ratio

Studying the power spectral density of a signal as in (2.24), but one that has been subjected to nonlinear processing at the transmitter, done in Fig. 3.1, it

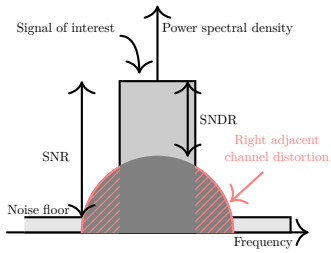


Figure 3.1: Illustration of signal-to-noise ratio (SNR), signal-to-noise and distortion ratio (SNDR) and how spectral broadening yields distortion power in the adjacent channels.

is seen that nonlinearities lead to so-called spectral broadening. That means the distortion spreads from the transmit band, to adjacent bands. To quantify how large this spread is, the adjacent channel leakage ratio (ACLR) can be computed. For a continuous signal the power in-band, versus the power of an adjacent band, is computed with an integral over the power spectral density in different frequency regions. In a discrete approximation, this is instead done with summations. Let $\hat{\boldsymbol{x}}$ be the frequency domain representation of $\boldsymbol{x}$. Then, the ACLR can be computed as

$$\text{ACLR} = \frac{\sum_{k\in\kappa_{\text{ib}}} |\hat{x}[k]|^2}{\sum_{k\in\kappa_{\text{adj}}} |\hat{x}[k]|^2}\,, \tag{3.12}$$

where k is a frequency index, κ_{ib} denotes in-band indices and κ_{adj} denotes indices in an adjacent band. This metric is most relevant to quantify in a single-PA transmitter, at the PA output, as it will then also directly reflect the radiated out-of-band distortion power.

A practical concern computing (3.12) can be exactly which indices k to include in κ_{ib} and κ_{adj}. This since for a finite signal, the power of the input signal $\boldsymbol{x}$ might not be perfectly contained within the bandwidth $[-\text{BW}/2, \text{BW}/2]$, which can make the ACLR seem unreasonably high if computed exactly as indicated in Fig. 3.1. Including a guard band, i.e. computing the adjacent bands at $[\text{BW}/2 + \Delta, 3\text{BW}/2 + \Delta]$ and $[-3\text{BW}/2 - \Delta, -\text{BW}/2 - \Delta]$ is a way to make the metric more true to the actual emitted adjacent band distortion. Additionally, as per the assumptions in previous chapters, filtering is performed after the PA and thus included in the digital signal representation.

To fulfill emission regulations for commercial applications, 3GPP specifies the type of filter, Δ, and procedure for ACLR computations as in (3.12).

Normalized mean square error

The normalized mean square error (NMSE) quantifies the residual error comparing two discrete signal vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ as

$$\text{NMSE} = \frac{\sum_{n=1}^{N}(y[n] - x[n])^2}{\sum_{n=1}^{N} y[n]^2} \,. \tag{3.13}$$

Unlike ACLR, it makes no distinction if the error is in-band or out-of-band. Unlike SNDR, it states no explicit distinction between what effect causes the residual error, be it noise, distortion or some other impairment. Finally, to compute the NMSE correctly, careful time-alignment must be performed to adjust for the case when $\boldsymbol{y}$ is delayed relative to $\boldsymbol{x}$.

CHAPTER 4

Linearization using single-input, single-output digital predistortion

In Chapter 2, a simplified overview of how a digital signal x can be propagated from a transmitter, through a channel, and to a receiver was shown. In this chapter, with the extended knowledge of the impairments induced by the PA from Chapter 3, linearization using DPD will be detailed. The general goal throughout this chapter is to study how one can obtain a desired, linear output, from an input signal passed through a DPD and a PA.

4.1 Role of digital predistortion

Building upon knowledge of PA nonlinearity, the DPD serves the purpose of processing the signal to invert the PA's nonlinear effect upon it. That is, if the PA exhibits compressive nonlinear behaviour, the DPD will be expansive, in addition to compensating for induced memory effects. For this section, the governing equation for the generated output y will be on the generic form

$$y = f_{\mathrm{PA}}(f_{\mathrm{DPD}}(x)) \,, \tag{4.1}$$

for some input x, and the purpose of the DPD will be to minimize the deviation from a desired output y_{d}, commonly measured as

$$\min ||y - y_{\mathrm{d}}||^2 . \tag{4.2}$$

This is equivalent to minimizing the NMSE as in (3.13). In implementation, this has historically been done mostly using time-domain modeling, which will also be the initial focus, with a comment on frequency-domain implementation in a later subsection.

It should be noted that in (4.1), f_{DPD} should naturally be a function applied in the digital domain. On the other hand, $f_{\mathrm{PA}}(\cdot)$ represents a function applied in the analog domain at RF. However, as was be detailed in Section 3.4, an equivalent baseband model of the PA can be made. As such, both these functions will for the remainder of this chapter be considered in discrete time, at baseband.

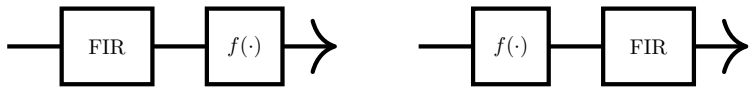
4.2 Invertability of nonlinear PAs

A fundamental concept for time-domain DPD modeling is the so-called theory of p:th order inverses of nonlinear systems. Define the joint nonlinear system $Q = f_{\mathrm{PA}}(f_{\mathrm{DPD}})$, and let Q_p be defined as the Volterra-series terms modeling that system, for the nonlinear terms of order p. Then it was shown in [20] that a nonlinear function f_{DPD} of order P can be found such that (4.1) becomes

$$y[n] = x[n] + \sum_{p=P+1}^{\infty} Q_p(x[n]) . \tag{4.3}$$

That is, all lower order nonlinearities $p = 2, .., P$ can be cancelled with an appropriate choice of f_{DPD}, though higher order terms are injected. This result holds both if f_{DPD} is applied prior to, or after, f_{PA}, referred to respectively as a pre-inverse or post-inverse. In fact, [20] shows that if characterizing f_{DPD} as a post-inverse, it can then also be used as an approximate pre-inverse.

Letting $P \to \infty$, the system can in theory be made completely linear. In practice, it is neither feasible to let $P \to \infty$, nor is it necessarily the best strategy for a finite P to adopt f_{DPD} such that (4.3) holds. This since the residual distortion remaining in the high order terms $p > P$ might be reduced if some residual distortion is allowed in lower order terms $p \leq P$. The theory



(a) Wiener. **(b)** Hammerstein.

Figure 4.1: Schematic for Wiener and Hammerstein architectures.

and observations leave two topics to discuss. Firstly, what DPD model to adopt for f_{DPD} to yield minimal residual distortion. Secondly, how to find the parameters of that DPD model, which is referred to as DPD identification. These topics will be discussed in the coming sections.

4.3 Digital predistortion models

Digital predistortion modeling is a wide topic, with many architectures and adaptations proposed throughout the years. Some common ones include Wiener, Hammerstein, look-up tables, piecewise models such as the decomposed vector rotation-DPD, and various types of pruned and truncated Volterra-series. In the following subsections, these will briefly be detailed.

Wiener and Hammerstein models

Wiener models for DPD, seen in Fig. 4.1a, are characterized by a linear filter of the input vector $\boldsymbol{x}$ as

$$z[n] = \alpha_0 x[n] + ... + \alpha_M x[n - M]\,, \tag{4.4}$$

to consider the memory effects of the PA, followed by a memoryless nonlinear function to compensate for the PA's static nonlinearity

$$u[n] = f(z[n])\,. \tag{4.5}$$

Augmented architectures include, for example, variations on the filtering process prior to the nonlinearity [21].

Hammerstein models, in contrast, place the nonlinearity prior to the linear filtering, seen in Fig. 4.1b.

Look-up tables

The use of look-up tables is a data-driven approach to DPD, where the corresponding predistorted data value $y[n]$ for a given input data value $x[n]$ is found via tabular look-up. Exemplified, a tabular look-up with memory can be formulated as

$$y[n] = \sum_{m=0}^{M} x[n-m] f_{\mathrm{LUT},m}(|x[n-m]|)\,. \tag{4.6}$$

In the simplest form, $f_{\mathrm{LUT},m}(|x|)$ is a pre-learned look-up table of scalar multiplicative factors based on the input magnitude of $|x|$. However, $f_{\mathrm{LUT},m}(|x|)$ can for example also incorporate interpolation and extrapolation [22], and recursive architectures [23] instead of (4.6) have been proposed, to reduce the tabular size needed for look-up.

Decomposed vector rotation-based DPD

The decomposed vector rotation (DVR) DPD model nonlinearities and memory effects through piecewise vector decomposition. The model is in [24] formulated[1] to include terms such as

$$\begin{aligned} y[n] = &\sum_{m=0}^{M} a_m x[n-m] \\ &+ \sum_{i=1}^{I}\sum_{m=0}^{M} c_{im,1}||x[n-m]| - \beta_i| e^{j\phi[n-m]} \\ &+ \sum_{i=1}^{I}\sum_{m=0}^{M} c_{im,21}||x[n-m]| - \beta_i| e^{j\phi[n-m]}|x[n]| \\ &+ \ldots \end{aligned} \tag{4.7}$$

Here, nonlinear terms are generated through absolute value operators, capable of generating high order terms with respect to x. This has been shown effective in being able to linearize highly nonlinear PAs with few terms [24], and possible to be implemented efficiently in hardware [25].

[1] rewritten to be consistent with this thesis' notation

Volterra-based DPD models

Volterra-series based time-domain DPD have been thoroughly researched, in particular on the topic of model reduction, especially in accounting for memory terms. This has yielded a few common pruned models of particular interest. The memory polynomial [26] is one simple adaptation, but one with many extensions and variations on e.g. pruning, frequency selectivity and envelope adaptation [27]–[30]. In a general form it is stated as

$$u[n] = \sum_{m=0}^{M} \sum_{p=1}^{P} c_{p,m} x[n-m]|x[n-m]|^{p-1}\,, \tag{4.8}$$

accounting for nonlinearities and memory effects without any cross terms. The generalized memory polynomial [31] can be stated as

$$\begin{aligned} u[n] = &\sum_{m=0}^{M} \sum_{p=1}^{P} c_{p,m} x[n-m]|x[n-m]|^{p-1} \\ &+ \sum_{m=0}^{M} \sum_{g=-G}^{G} \sum_{p=1}^{P} c_{p,m,g} x[n-m]|x[n-m-g]|^{p-1}\,, \end{aligned} \tag{4.9}$$

and accounts for some of the memory cross-terms of the full Volterra-series that are often dominant.

Extensive research has been performed on the topic of optimally choosing which bases to include from the GMP or full Volterra-based DPD models [32]–[35].

Frequency-domain digital predistortion

In contrast to time-domain DPD, and as the name would indicate, frequency-domain (FD) DPD is applied on frequency-domain data. FD-DPD can be motivated, as in e.g. [36] when DPD is only necessary to apply over a subset of frequencies. However, time-domain, bandlimited DPDs as in [37] can also address the issue of only linearizing a subset of frequencies. The complexity of implementation will vary though, and is as such a determining factor.

4.4 Digital predistortion identification

With a given DPD model chosen, one practical challenge is identification, i.e. finding model coefficients. For example, which coefficients $c_{p,m}$ should be chosen in (4.8). Identification procedures can generally be categorized into direct learning architectures, and indirect learning architectures. These will be detailed below.

In direct learning architectures, the optimization procedure directly changes the DPD model coefficients in the search for an optimal set. One example of such a method is the Gauss-Newton algorithm. Given input data $\boldsymbol{x}$, output data $\boldsymbol{y}$, and a model $f(\boldsymbol{x}, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are the model parameters, Gauss-Newton iterations require the computation of the residuals

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{y} - f(\boldsymbol{x}, \boldsymbol{\theta})\,. \tag{4.10}$$

Denoting r_i as the i:th sample of $\boldsymbol{r}$, the Jacobian matrix can be formed as

$$(\boldsymbol{J})_{i,j} = \frac{\partial r_i(\boldsymbol{\theta}^{[q]})}{\partial \boldsymbol{\theta}_j}\,, \tag{4.11}$$

where $[q]$ denotes an iteration index. With these, the iterations are computed as

$$\boldsymbol{\theta}^{[q+1]} = \boldsymbol{\theta}^{[q]} - (\boldsymbol{J})^{-1}_{i,j}\boldsymbol{r}\,. \tag{4.12}$$

Practically, computing (4.10) to (4.12) is nontrivial, since $f(\boldsymbol{x}, \boldsymbol{\theta})$ is a nonlinear function composed of the tandem operation of a nonlinear DPD followed by a nonlinear PA. As such, approximate and modified forms of (4.12) are found in many works, e.g. in [38]–[40].

Indirect learning architectures, in contrast, do not directly adjust the DPD modeling coefficients each iteration. One kind of indirect learning utilizes the p:th order inverse theory to identify a post-inverse, which is then applied as a pre-inverse. In this, a specific DPD model is assumed. Exemplified, assume $\boldsymbol{y}_\mathrm{d}$ is an input to some PA, resulting in a PA output $\boldsymbol{y}$. Let the post-inverse model basis be based on (4.8), yielding

$$\boldsymbol{F}(\boldsymbol{y}) = \begin{bmatrix} \boldsymbol{y}[1], \ldots\ \boldsymbol{y}[1]|\boldsymbol{y}[1]|^{P-1}, & 0 & \ldots & 0 & \ldots \\ \boldsymbol{y}[2], \ldots\ \boldsymbol{y}[2]|\boldsymbol{y}[2]|^{P-1}, & \boldsymbol{y}[1] & \ldots\ \boldsymbol{y}[1]|\boldsymbol{y}[1]|^{P-1} & \ldots \end{bmatrix} \tag{4.13}$$

If the desired output of the PA is $\boldsymbol{y}_\mathrm{d}$, the model coefficients $c_{p,m}$ are then identified as

$$\boldsymbol{c} = \begin{bmatrix} c_{1,0} \\ c_{1,1} \\ \vdots \\ c_{P,M} \end{bmatrix} = \boldsymbol{F}(\boldsymbol{y})^\dagger \boldsymbol{y}_\mathrm{d}\,. \tag{4.14}$$

Here, $\boldsymbol{F}(\boldsymbol{y})^{\dagger}$ denotes the pseudo-inverse of $\boldsymbol{F}(\boldsymbol{y})$. Using this model as a pre-inverse, for predistortion, one simply adapts the input to (4.13) to be $\boldsymbol{y}_{\mathrm{d}}$ instead of $\boldsymbol{y}$ but then keep the modeling coefficients of (4.14). This approach was used in [41].

In contrast to this approach of indirect learning, a different kind of indirect learning architecture is the so-called iterative learning control (ILC) [42] which makes no assumption on the DPD model. Instead, the input signal is directly modified to linearize the output, by a scheme

$$\boldsymbol{x}^{[i+1]} = \boldsymbol{x}^{[i+1]} - \eta(\boldsymbol{y}_{\mathrm{d}} - \boldsymbol{y})\,, \tag{4.15}$$

where η is a scalar step length, $\boldsymbol{y}$ is the measured PA output and $\boldsymbol{y}_{\mathrm{d}}$ the desired PA output. This can be used as a reference for how well a PA is linearizable without DPD model assumptions. Alternatively, upon ILC convergence, $\boldsymbol{x}^{[i+1]}$ can be used as a target signal for mapping from the original input signal via some DPD model.

CHAPTER 5

Digital predistortion in multiple-input, multiple-output settings

With multiple input signals, and multiple output signals, passed through a transmitter the system architecture typically grows in size and complexity. Balancing the positive benefits of such a system, e.g. beamforming capabilities in a MIMO transmitter or capability to transmit at multiple frequencies to different users concurrently with the same PA, to the added system complexity is a non-trivial challenge with trade-offs. In this chapter, the role of DPD in such systems is considered. In particular, the first section treats considerations for the multi-band setting. The second section follows up by describing different architectures for concurrent multi-beam transmission in MIMO transmitters. Lastly, the challenges and opportunities for characterization of DPDs using over-the-air observations are considered.

5.1 Multi-band digital predistortion

Transmitting data concurrently to users across different bands, through the same PA, allows a radio basestation to increase its capacity significantly. If

these bands are adjacent, the total bandwidth can be considered a single wideband input signal to be linearized. However, if they are far spaced in frequency, the problem is sparse and wideband linearization perhaps unnecessarily computationally expensive. Sparsity can be necessary, as previously mentioned, due to spectrum congestion around certain frequencies, but also due to that some users might require different carriers for favorable propagation conditions and acceptable SNR.

To understand the operation of a multi-band DPD, the nonlinear terms induced by multi-band transmission must be understood. This will be presented in the first subsection, followed by listing a few multi-band DPD approaches.

Multi-band transmission through nonlinear system

Consider first the case of transmitting a dual-band signal through a wideband transmitter. Let x_1 denote data to a user to be transmitted at frequency f_1, and x_2 data to a user at f_2. Let their respective bandwidths be BW_1 and BW_2. Transmitting these concurrently yields

$$x[n] = x_1[n]e^{j2\pi f_1 n/f_s} + x_2[n]e^{j2\pi f_2 n/f_s}\,. \tag{5.1}$$

Assume this digital signal goes through the transmitter processing, a channel and then a receiver as in Chapter 2. Let the PA be represented with a baseband equivalent model, as detailed in Section 3.4, though for analytical brevity simplified such that

$$f_{\mathrm{PA}}(x[n]) = c_1 x[n] + c_3 x[n]|x[n]|^2\,. \tag{5.2}$$

Passing the signal of (5.1) through the nonlinearity of (5.2), the output can be expanded as

$$\begin{aligned}
y[n] &= c_1(x_1[n]e^{j2\pi f_1 n/f_s} + x_2[n]e^{j2\pi f_2 n/f_s}) \\
&\quad + c_3(x_1[n]e^{j2\pi f_1 n/f_s} + x_2[n]e^{j2\pi f_2 n/f_s}) \cdot |x_1[n]e^{j2\pi f_1 n/f_s} + x_2[n]e^{j2\pi f_2 n/f_s}|^2 \\
&= c_1(x_1[n]e^{j2\pi f_1 n/f_s} + x_2[n]e^{j2\pi f_2 n/f_s}) \\
&\quad + c_3(x_1[n](|x_1[n]|^2 + |x_2[n]|^2)e^{j2\pi f_1 n/f_s} \\
&\qquad + x_2[n](|x_1[n]|^2 + |x_2[n]|^2)e^{j2\pi f_2 n/f_s} \\
&\qquad + x_1[n]^2 x_2[n]^* e^{j2\pi(2f_1 - f_2)n/f_s} \\
&\qquad + x_2[n]^2 x_1[n]^* e^{j2\pi(2f_2 - f_1)n/f_s})\,.
\end{aligned} \tag{5.3}$$

Studying (5.3), two types of intermodulation (IM) products are noted. Firstly, the type $x_i|x_j|^2$ that always ends up at frequency f_i, $i, j = 1, 2$. Secondly, the type $x_i[n]^2 x_j[n]^*$ that end up at frequency $2f_i - f_j$, $i, j = 1, 2, i \neq j$, i.e. always away from the carrier frequency unless f_s is small enough to yield aliasing. Adding a third term $x_3[n]e^{j2\pi f_3 n/f_s}$ in (5.1), and expanding via (5.2), a third type of IM products of the form $x_{i_1}[n]x_{i_2}[n]x_{i_3}^*[n]$ that end up at frequency $f_{i_1} + f_{i_2} - f_{i_3}$ will be induced. This type of IM categorization was done in previous work such as [43], [44], with respective notation as Type I, Type II and Type III. These nonlinear products must be accounted for, and often compensated. In the following subsection, multi-dimensional DPD methods will be discussed.

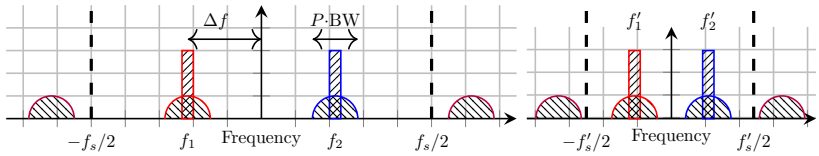
Multi-dimensional digital-predistortion methods

In [45], a dual-input, dual-output DPD was proposed for addressing concurrent dual-band linearization. A more general DPD was then proposed in [46], based on a full Volterra series, which was shown possible to reduce to the model of [45]. In essence, [46] details inputting (5.1) into (3.6) and expanding, followed by proposing various ways of complexity reduction. In addition to (5.3), a full expansion yields higher order terms, more cross terms and includes memory effects for the concurrent dual-band case.

In [47], concurrent tri-beam linearization is addressed. It then proposes generating the streams

$$\begin{aligned} y_1 &= \sum_m \sum_p \sum_s \sum_i x_1[n-m]|x_1[n-m]|^{p-s}|x_2[n-m]|^{s-i}|x_3[n-m]|^i\,, \\ y_2 &= \sum_m \sum_p \sum_s \sum_i x_2[n-m]|x_1[n-m]|^{p-s}|x_2[n-m]|^{s-i}|x_3[n-m]|^i\,, \\ y_3 &= \sum_m \sum_p \sum_s \sum_i x_3[n-m]|x_1[n-m]|^{p-s}|x_2[n-m]|^{s-i}|x_3[n-m]|^i\,, \end{aligned} \tag{5.4}$$

to be placed at the corresponding three center frequencies f_1, f_2, f_3. Even without adding all cross-terms, as a full expansion in the style of [46] would, it becomes evident that the number of terms necessary to generate increases sharply with the increased number of bands. The formulation (5.4) considers only intermodulation products occuring at the fundamental frequencies transmitted. However, as illustrated in (5.3), concurrent multi-band transmission



(a) Sparse dual-band signal. **(b)** Relocated dual-band signal.

Figure 5.1: Power spectral density of dual-band signals passed through nonlinearity.

yields cross-terms at other frequencies as well. These might, or might not be, filtered by analog bandpass filters depending on the band spacing. In [44], a prefrequency shift technique was introduced to account for Type I and Type II IM products, whilst all types are addressed for closely spaced multi-band signals in [48].

Many of the listed approaches so far can be extended to include more bands. However, with all of these works being multi-input, multi-output DPDs, the dimensionality and complexity will increase rapidly. This has motivated works looking into if single-input, single-output DPDs could address the multi-band setting.

Frequency relocation

An alternative approach to multi-band linearization involves so-called frequency relocation [49]. Revisiting (5.3) it is realized that in order to sample this signal accounting for the main signal components and their carrier-centered spectral regrowth, then

$$f_s \geq 2\Delta f + P \cdot \mathrm{BW}\,, \tag{5.5}$$

where $\Delta f = |f_2 - f_1|/2$, visualized in Fig. 5.1a. If IM terms are induced beyond this frequency range, they are aliased back into the range during sampling. However, it appears from Fig. 5.1a that the spectrum resulting from having bands placed at f_1, f_2 could be densified. Keeping the relative frequency spacing between the bands, but relocating bands to lie closer to one another before a nonlinear function, all induced nonlinear products will occur

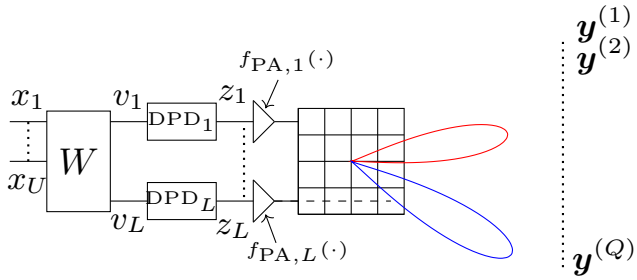


Figure 5.2: Per-PA DPD in MIMO transmitter.

at the same relative positions. However, the sampling can then instead be performed at

$$f_s' = 2\Delta f' + P \cdot \mathrm{BW}\,, \tag{5.6}$$

where $\Delta f' = |f_2' - f_1'|/2$ is the band spacing after relocation. This is seen in Fig. 5.1b. Afterwards, bandpass filtering and inverse relocation of bands can be performed to obtain the original signal of Fig. 5.1a.

One practical consideration using frequency relocation is the variations in PA memory effects at different RF frequencies, as mentioned in Chapter 3. Left unaccounted for, one might concatenate two signals with very different memory responses when performing frequency relocation. This in turn would lead to a DPD requiring many memory terms to compensate for the PA variations. This was addressed in Paper A, where bandwise filters are implemented prior to the DPD to equalize the PA response.

5.2 Multi-beam digital predistortion

With multiple transmit chains and antennae, the ability to beamform has yielded a new setting for DPD. Conventional linearization is to apply one DPD per PA, visualized in Fig. 5.2. U user data streams $x_1, ..., x_U$ are precoded to L transmit branches as signals v_l for $l = 1, ..., L$, where DPD is applied on each branch before the signal passes through PAs, antennae and a channel, to be observed in Q directions. Performing computations along the antenna branches will be referred to in the following text as doing DPD in the antenna space. With per-PA DPD in the antenna space, complexity scales linearly with L.

Challenging per-PA DPD, a highly relevant setting is when the transmitter has many branches L in relation to the number of users U. The mathematics of the problem is in many ways similar to that of the multi-band problem, and in some ways different. Revisiting (2.13), observe two data streams x_1, x_2 on branch l that are beamformed with weights $w_{l,1}, w_{l,2}$ as

$$z_l = w_{l,1}x_1 + w_{l,2}x_2\,, \tag{5.7}$$

and passed through a third order nonlinearity $f(x) = c_1c_3x + x|x|^2$ such that

$$\begin{aligned} f(z_l) =& c_1z_l + c_3z_l|z_l|^2 \;= \\ =& c_1w_{l,1}x_1 + c_1w_{l,2}x_2 \\ &+c_3(w_{l,1}x_1 + w_{l,2}x_2)\cdot|w_{l,1}x_1 + w_{l,2}x_2|^2 \\ =& c_1w_{l,1}x_1 + c_3w_{l,1}x_1(|x_1|^2 + |x_2|^2) \\ & c_1w_{l,2}x_2 + c_3w_{l,2}x_2(|x_1|^2 + |x_2|^2) \\ &+c_3w_{l,1}^2w_{l,2}^*x_1^2x_2^* \\ &+c_3w_{l,2}^2w_{l,1}^*x_2^2x_1^*\,. \end{aligned} \tag{5.8}$$

The similarities to (5.3) are clear in how some intermodulation terms will be induced in the same directions as the linear user data streams, whilst other occur elsewhere. This follows from beamforming theory, briefly detailed in subsection 2.2. The PA outputs will either add up more constructively, or destructively, in different spatial direction when propagated through the channel depending on the beamforming weights $w_{l,1}, w_{l,2}$. The last two rows of (5.8) correspond to terms in the so-called off-beams. In analogy to (5.3), the channel or antennae spatial gain pattern might functionally operate to attenuate off-beams, as the bandpass filter can do for IM terms in the multi-band setting. However, this will not always be the case.

In the example above, regardless of L, there are only $B = 4$ unique streams of data going in different spatial directions. These streams are the original data streams plus distortion, $x_1 + x_1(|x_1|^2 + |x_2|^2)$ and $x_2 + x_2(|x_1|^2 + |x_2|^2)$, and the generated IM streams in new off-beam directions $x_1^2x_2^*$ and $x_2^2x_1^*$. Thus, for $L \gg B$, the system is sparse. This has motivated DPD approaches that operate on the streams $x_1, ..., x_U$ and their nonlinear products, instead of operating on the precoded streams v_l. These approaches will be referred to as operating in the beam domain. It is important to note that whilst per-PA DPD, as seen in Fig. 5.2, scales in complexity linearly with L, any approach

Table 5.1: Off-beam terms for static non-linearity of order $P = 3$, and with $U = 2, 3, 4$ concurrent users.

Number of users	Off-beam terms	Number of off-beam terms
$U = 2$	$x_1x_1x_2^*$, $x_2x_2x_1^*$	2
$U = 3$	$x_1x_1x_2^*$, $x_2x_2x_1^*$, $x_1x_1x_3^*$, $x_3x_3x_1^*$ $x_2x_2x_3^*$, $x_3x_3x_2^*$, $x_1x_2x_3^*$, $x_1x_3x_2^*$, $x_2x_3x_1^*$	9
$U = 4$	$x_1x_1x_2^*$, $x_2x_2x_1^*$, $x_1x_1x_3^*$, $x_3x_3x_1^*$ $x_2x_2x_3^*$, $x_3x_3x_2^*$, $x_1x_2x_3^*$, $x_1x_3x_2^*$, $x_2x_3x_1^*$, $x_1^2x_4^*$, $x_4^2x_1^*$, $x_2^2x_4^*$, $x_4^2x_2^*$, $x_3^2x_4^*$, $x_4^2x_3^*$, $x_1x_2x_4^*$, $x_1x_4x_2^*$, $x_2x_4x_1^*$, $x_1x_3x_4^*$, $x_1x_4x_3^*$, $x_3x_4x_1^*$, $x_2x_3x_4^*$, $x_2x_4x_3^*$, $x_3x_4x_2^*$	24

operating in the beam domain will scale in relation to the number of users and the induced number of intermodulation terms. This will be understood in greater detail in the next subsection.

Multinomial theory to determine beam domain sparsity

The number of intermodulation terms in beam domain approaches follows from multinomial theory. For a P:th order nonlinearity with U users, stated as

$$f(u) = (x_1 + ...x_U)|x_1 + ...x_U|^{P-1}, \tag{5.9}$$

the number of terms in $|x_1 + ...x_U|^{P-1}$ are

$$\binom{P-1+U-1}{U-1}, \tag{5.10}$$

whilst $(x_1 + ...x_U)$ contains U terms. As such, (5.9) generates

$$N_{\mathrm{IM},U,P} = U\binom{P-1+U-1}{U-1}, \tag{5.11}$$

P:th order terms. Compared to (5.8), $P = 3$, $U = 2$ in (5.11) indeed generates $2 \cdot 3 = 6$ third order terms, namely $x_1|x_1|^2, x_1|x_2|^2, x_2|x_1|^2, x_2|x_2|^2, x_1^2x_2^*, x_2^2x_1^*$. However, considering $P = 3$, $U = 3, 4, 5$ then the number of IM terms are 18, 40 and 75, respectively. An important note to this is that not all of these

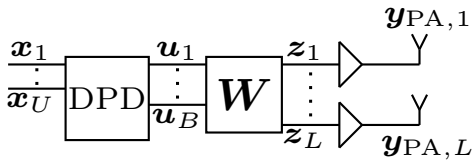


Figure 5.3: Beam-domain DPD for U users, precoded to generate B streams unique in space, transmitted over L transmit branches.

terms will be emitted in unique spatial directions, as is clear from (5.8). In fact, many of these will be emitted in user directions. Table 5.1 details the off-beam terms for a third order nonlinearity, and $U = 2, 3, 4$ users, to showcase this.

To linearize a system as in (5.8), two approaches will be detailed in the following subsections. Firstly, so-called beam-domain DPD, and secondly a novel contribution from referred to as virtual-array DPD.

Beam-domain digital predistortion

The works of [50] and [51] provide two examples of beam-domain DPD. These approaches operate similarly to how 2D-DPD and 3D-DPD did in the multi-band setting, namely by a multi-dimensional DPD operating on the input data stream to generate appropriate multidimensional predistorted streams. Beam-domain DPD is schematically visualized in Fig. 5.3, where B denotes the number of unique main beams and off-beams. To compensate for the PA example of (5.8), $U = 2$ and $B = 4$, with this scheme, time-domain streams

$$\begin{aligned}
\boldsymbol{u}_1 &= \theta_1 x_1 + \theta_3 x_1(|x_1|^2 + |x_2|^2) \\
\boldsymbol{u}_2 &= \theta_1 x_2 + \theta_3 x_2(|x_1|^2 + |x_2|^2) \\
\boldsymbol{u}_3 &= \theta_3 x_1^1 x_2^* \\
\boldsymbol{u}_4 &= \theta_3 x_2^1 x_1^* ,
\end{aligned} \tag{5.12}$$

can be generated by the multi-input, multi-output DPD, where θ_1, θ_3 are DPD model coefficients. Implicit in generating these predistorted signals is that they are to linearize PAs that are, at least approximately, the same across different branches. Seen in Fig. 5.3, the DPD is applied prior to beamforming, and the DPD coefficients can not at a stream-level account

for PA variations on each branch. The DPD stream generation is followed by applying beamforming weights in $\boldsymbol{W}$ for these streams to go in the right spatial directions.

The downsides of beam-domain DPD are twofold. As a first downside, beam-domain DPD scales poorly with U, unless off-beams are neglected. Neglecting off-beams might be motivated under certain circumstances, and for relatively large U it was shown in [52] that the distortion tends to a uniform spread across space with increasing U. However, this does not necessarily hold that well for smaller U values. A second downside is the multidimensionality of bases in BD-DPD. For polynomial base implementation, the multiplications $x_1|x_1|^2$ and $x_2|x_1|^2$ might be equivalent to compute. However, for other DPD methods such as look-up-tables, multidimensional bases are more computationally expensive [53]. This motivates looking into alternative DPD methods for this setting.

Virtual array digital predistortion

From this chapter so far, two observations should be stated clearly. Firstly, if the number of antennae L is approximately the same as the number of beams B, then neither the antenna domain, nor the beam-domain, will be sparse. Per-PA DPD is then a computationally effective single-input, single-output linearization solution. Secondly, when U increases, the dimensionality of the multi-input, multi-output DPD of Fig. 5.3 increases undesirably. Putting these observations into the context where the number of antennae is far greater than the number of unique data streams, $L \gg B$, this section will show a novel contribution on how linearization can be done in a virtual array of size $L_{\text{v}} \approx B$ which utilizes the upside of the first observation, whilst avoiding the short coming the second observation. This work is later found in greater detail in Paper B.

In Fig. 5.4, the proposed virtual array (VA) DPD is shown. The VA-DPD aims to produce predistorted user streams and off-beam IM terms in a low-dimensional virtual array, and use these to linearize the larger, real transmit array. This relies on carefully considering where all predistorted terms end up in the virtual domain, such that they after the virtual array processing can be used to linearize the real array. In detail, U user data streams $x_1, ..., x_U$ are beamformed in a virtual domain of size L_{v}. In particular, discrete Fourier transform (DFT) based beamforming is performed, implemented with Fast

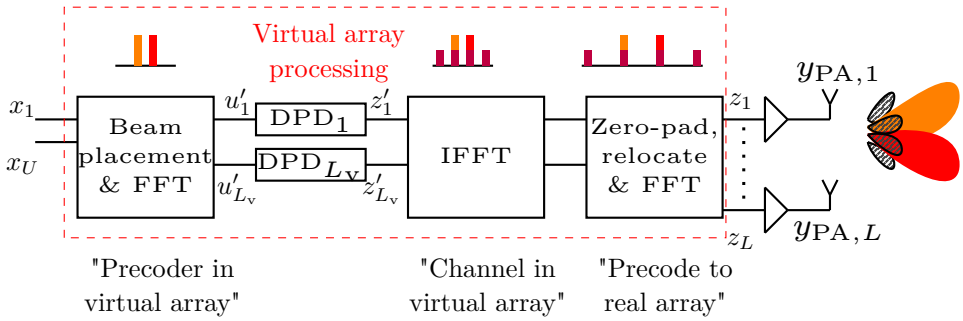
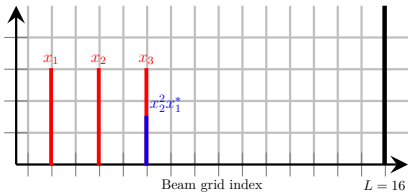


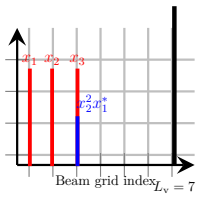
Figure 5.4: Schematic of DPD using virtual array processing. The bar plots exemplify the beam space representation of beamforming the data streams to beams and induced IM beams. Red and orange correspond to desired user signals, purple to the injected distortion by the DPDs to cancel out the unwanted distortion, shown in striped dark gray.

Fourier transforms (FFTs), placing user streams at indices $b'_1, ..., b'_U$ on a grid of beams. After the DPD nonlinearities, a set of $\{b'_1, ..., b'_U, ..., b'_B\}$ beam directions will exist with corresponding data streams. These data streams are separable, leading to that the streams can be arranged correctly for precoding onto the real array. This is exemplified in Fig. 5.5. Here, an example is provided where after the DPD, FFT-beamforming is performed to precode the user streams to beam indices $\{b_1, ..., b_3\} = \{2, 4, 6\}$ on a beamgrid corresponding to a uniform linear array of size $L = 16$. In Fig. 5.5a, the occurrence of main beams, and the off-beam corresponding to the IM term $x_2^2 x_1^*$, is seen. Here, it is noted that the stream $x_2^2 x_1^*$ will be emitted in the same spatial direction as x_3. Other IM terms, omitted from the illustration for clarity, will either end up on unique beam grid indices or overlapping with $\{b_1, b_2, b_3\}$. The full set of beam indices in the real array will be $\{b_1, b_2, ..., b_B\}$.

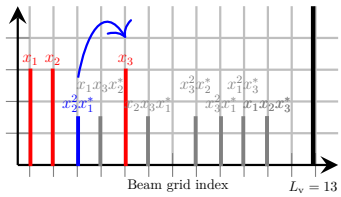
Fig. 5.5b and Fig. 5.5c illustrates two different virtual domains, of size $L_v = 7$ and 13, respectively, with different virtual beam placements on the grid. In Fig. 5.5b, $\{b'_1, b'_2, b'_3\} = \{1, 2, 3\}$ is used, and it is observed that in this virtual domain $x_2^2 x_1^*$ still ends up in the same direction as x_3. Thus, as far as the main beams and the IM term $x_2^2 x_1^*$ is concerned, the data representation in Fig. 5.5b can be transferred to that of 5.5a by simply padding the beam grid with zero entries and relocating data corresponding to beam indices. Should this hold for all other generated IM terms as well, it follows that the



(a) Beam occurrences in real array.



(b) $L_\mathrm{v} = 7$, $\{b'_1, b'_2, b'_3\} = \{1, 2, 3\}$.



(c) $L_\mathrm{v} = 13$, $\{b'_1, b'_2, b'_3\} = \{1, 2, 5\}$.

Figure 5.5: Beam occurrences for $\{b_1, b_2, b_3\} = \{2, 4, 6\}$ in real and virtual arrays, with placement of $x_2^2 x_1^*$ highlighted.

virtual domain $L_v = 7$ with $\{b'_1, b'_2, b'_3\} = \{1, 2, 3\}$ can represent the real array without loss. Finer mathematical detail to exactly how this can be performed is provided in Paper B. In addition, to complement Fig. 5.5b, Fig. 5.5c illustrates another virtual domain representation where all IM terms are explicitly shown. Here, the IM term $x_2^2 x_1^*$ does not immediately overlap with x_3, as it would in the real array. However, it ends up on the virtual beam grid at a unique location, i.e. a location with no other IM terms. Thus, it can be moved to the beam grid index corresponding to x_3, achieving the same result.

Fundamental differences from the domain reduction techniques used in previous multi-band works [49], [54] are twofold. Firstly, the dimensionality reduction is not done in the time or frequency domain. Spectral regrowth around carriers does, as such, not need to be considered and the possibility of partial spectral overlap is removed. Furthermore, the ability to model memory effects is not diminished. Secondly, the orthogonality of the DFT-beamgrid allows for easily discarding samples and zero-padding, as part of the filtering.

Through derivations and simulations in Paper B, it is shown that the virtual domain can be made such that $L_\mathrm{v} \approx B$, thus leading to large savings in computational complexity when $B \ll L$. Furthermore, in contrast to BD-

DPD, the DPDs of Fig. 5.4 are single-input, single-output. Thus, the virtual domain DPD will scale with B in terms of the number of DPDs needed, but will never increase the dimensionality of the DPD units. Furthermore, Paper B shows that this DPD scheme can be implemented with other precoding techniques as well, and that it is more robust towards PA variations than the beam-domain approach.

5.3 Digital predistortion using over-the-air observations

Traditionally, DPD characterization and evaluation of residual distortion has been done by observing the PA output signals directly. However, in MIMO transmitters with many branches L, this approach can be challenged by the proposition of using over-the-air measurements instead [55]–[57]. Firstly, from a practical standpoint, it can be infeasible to measure each PA output in a dense, large array. Secondly, measurements at the PA output may not optimally reflect the desired system performance. This first point is straightforward in that implementing output couplers at each PA output is not only costly, but can be difficult from a signal routing perspective. The second point will receive more detailed descriptions over the coming subsections.

Distortion shaping in space

Revisiting Fig. 5.2, it can be noted that measuring distortion at the output of each PA will disregard any effect that the antennae or the channel have on the emitted power. This and remaining subsections of this chapter, following Paper C, will argue that minimizing the distortion at each PA's output is not necessarily equivalent to obtaining the best linearization performance over-the-air.

Consider an array consisting of L branches, as in Fig. 5.6. As defined in Chapter 4, for SISO DPD, the goal for linearization when measuring at the PA output typically follows from (4.2). That is, the deviation of the PA output $y_{\mathrm{PA},l} = f_{\mathrm{PA},l}(z_l)$, on branch l, from a desired output $y_{\mathrm{PA,d},l}$ should be minimized as

$$\min ||y_{\mathrm{PA},l} - y_{\mathrm{PA,d},l}||^2 . \tag{5.13}$$

This is equivalent to minimizing the NMSE, as in (3.13). Let, as in Fig. 5.6,

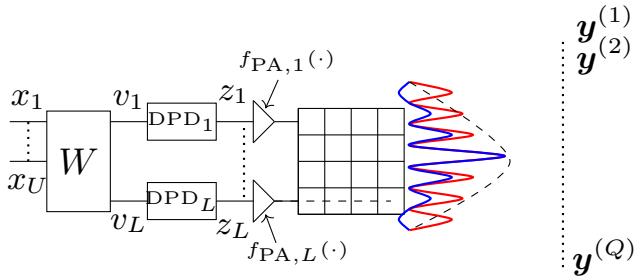


Figure 5.6: MIMO transmitter with spatial filtering due to narrow field-of-vision antennae. In red is the emitted power if all antennae radiated equally well in the angular range $(0, 180)^\circ$, whilst blue is the emitted power given the antenna spatial filter indicated with the dashed black line.

each branch l be equipped with a DPD with correponding coefficients $\boldsymbol{\theta}_l$, that will affect the PA output $y_{\mathrm{PA},l}$ on that branch. Then (5.13) can be rephrased in order to find the optimal DPD as

$$\arg\min_{\boldsymbol{\theta}_l} ||y_{\mathrm{PA},l}(\boldsymbol{\theta}_l) - y_{\mathrm{PA,d},l}||^2 \,, \tag{5.14}$$

for each branch $l = 1, ..., L$. Given this optimization critera, then if each PA and desired output is the same, $\boldsymbol{\theta}_1 = ... = \boldsymbol{\theta}_L$ will follow. The following paragraph will showcase an example where this is argued to be suboptimal from a full system perspective.

Let a transmitter be equipped with multiple antennae that each have a limited field-of-vision, radiating well to broadside but poorly in the endfire direction. The combined array antenna radiation pattern is graphically visualized by the dashed line in Fig. 5.6, and in relation to (2.16) represents an antenna array pattern with the radiative properties

$$a_{\phi,f} = \sin^3(\phi) \,. \tag{5.15}$$

For a considered narrow bandwidth, the frequency f can be omitted from consideration. The model (5.15) is a bit arbitrarily chosen, but loosely based on conducted over-the-air measurements.

Worth comment in regards to (5.15) is that power intended to radiate outside the field-of-vision is instead reflected back towards the transmitter, which

in practice often can be handled with a circulator, after which the power eventually dissipates into heat. Noting that much of the power and distortion generated at the PA output is actually reflected if they are emitted in the endfire directions, the question is if that distortion is equally important to compensate for in the DPDs as the distortion emitted towards the broadside direction. Dissipating power into heat in the transmitter due to antenna reflections is undesired, but unavoidable. Furthermore, the distortion power is typically tens of dBs less than the linear power, and as such has close to negligible impact on heat dissipation. In contrast, it can have a notable impact if radiated towards users or interfering with other operators over-the-air. This argues for that shaping the distortion spatially could actually be beneficial for overall system performance. Thus, consider the optimization problem

$$\arg\min_{\boldsymbol{\theta}} ||\boldsymbol{H}\boldsymbol{y}_{\mathrm{PA}}(\boldsymbol{\theta}) - \boldsymbol{y}_{\mathrm{d}}||\,, \tag{5.16}$$

in contrast to (5.14), where

$$\boldsymbol{y}_{\mathrm{PA}}(\boldsymbol{\theta}) = \begin{bmatrix} y_{\mathrm{PA},1}(\boldsymbol{\theta}_1) \\ \vdots \\ y_{\mathrm{PA},L}(\boldsymbol{\theta}_L) \end{bmatrix}, \; \boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_1 \; ... \; \boldsymbol{h}_L \end{bmatrix}, \tag{5.17}$$

and $\boldsymbol{y}_{\mathrm{d}}$ is the desired signal transmitted in directions $q = 1, ..., Q$, compared to Figure 5.6. Here, each vector $\boldsymbol{h}_l$ denotes the channel from branch l to the Q observation locations over-the-air. With the antenna pattern (5.15) as part of the channel, some of the elements in $\boldsymbol{h}_l$ will have lower magnitude, i.e. less channel gain. Thus, those directions are less impactful in (5.16). Increasing distortion in such directions, to the benefit of lower distortion in high antenna gain regions, can then decrease the summed error $||\boldsymbol{H}\boldsymbol{y}_{\mathrm{PA}}(\boldsymbol{\theta}) - \boldsymbol{y}_{\mathrm{d}}||$. This argues for the potential benefit of spatial distortion shaping.

A conceptually similar idea to (5.16) was proposed in previous work [58]. There, a common DPD alongside LUTs to control PA biasing in a hybrid MIMO array, were used to supress sidelobes as well as reduce distortion in the main beam direction. However, no explicit optimization criterion such as (5.16) is stated to consider distortion across different spatial directions.

Before reaching the metric proposed in Paper C, considerations in frequency will be discussed as well.

Distortion shaping in frequency

In-band distortion and out-of-band distortion are important to consider for different reasons during wireless transmission. In-band distortion will deteriorate system performance for the intended user, and any other adjacent system using the same frequencies. Out-of-band distortion needs to be regulated to not interfere with neighboring operators or systems, and is in SISO systems typically quantified with ACLR as defined in (3.12). In MIMO systems, it is also important to consider the spatial distribution of out-of-band radiation.

Lowering in-band distortion allows for a higher modulation order to be utilized for communication, at an acceptably low error rate. On the other hand, for lower modulation orders, the in-band distortion does not necessarily need to be very low for reliable communication. Instead, in-band distortion can be allowed to increase, in order to improve out-of-band distortion. One conventional way of performing this in practical communication systems is via various crest-factor reduction (CFR) methods. One common, iterative method is clipping-and-filtering [59]. Conceptually, the CFR reduces the PAPR, as defined in (3.10), whilst the DPD that typically follows afterwards increases PAPR. This, seemingly contradictory, process has not always been co-optimized to yield optimal linearization. Some works have however addressed joint CFR with DPD [60], [61], and others have addressed frequency-selective DPD approaches using time-domain filtering techniques [37], [62]. One recent paper also addresses the need for a frequency weighted approach to distortion compensation in DPDs, with a Pareto-optimal approach [63]. Lastly, in [36], a frequency-domain approach that allows for frequency selective linearization was proposed, allowing for flexible DPD linearization in the frequency domain.

Distortion shaping in both frequency and space

To the author's best knowledge, no previous work have addressed how the spatial dimension can be utilized in a MIMO setting to not only determine a good trade-off between in-band and out-of-band distortion, but also where to in space each type is to be emitted. As such, let each DPD, on branches $l = 1, ..., L$ be defined in a frequency domain formulation as

$$\hat{\boldsymbol{z}}_l[k] = \hat{\boldsymbol{F}}_l[k]\hat{\boldsymbol{\theta}}_l[k]\,, \tag{5.18}$$

where k is a frequency index, $\hat{\boldsymbol{z}}_l$ is the frequency domain representation of the DPD output $\boldsymbol{z}_l$, $\hat{\boldsymbol{F}}_l$ are DPD bases and $\hat{\boldsymbol{\theta}}_l$ DPD coefficients. For generating the vector of DPD inputs across all branches at frequency k, $\hat{\boldsymbol{z}}[k]$, then

$$\hat{\boldsymbol{z}}[k] = \hat{\boldsymbol{F}}[k]\boldsymbol{\Theta}[k] = \begin{bmatrix} \hat{\boldsymbol{F}}_1[k]\hat{\boldsymbol{\theta}}_1[k] \\ \hat{\boldsymbol{F}}_2[k]\hat{\boldsymbol{\theta}}_2[k] \\ \vdots \\ \hat{\boldsymbol{F}}_L[k]\hat{\boldsymbol{\theta}}_L[k] \end{bmatrix}. \tag{5.19}$$

Similarly, let $\hat{\boldsymbol{y}}_{\mathrm{d}}^{(q)}[k]$ be the frequency domain representation of the desired signal at spatial index q. Modifying (5.16), Paper C proposes the optimization criterion

$$\arg\min_{\boldsymbol{\Theta}} \sum_{q=1}^{Q} \sum_{k=1}^{K} w_{q,k} \left| \hat{\boldsymbol{y}}_{\mathrm{d}}^{(q)}[k] - \boldsymbol{h}^{(q)} f_{\mathrm{PA}}(\hat{\boldsymbol{F}}[k]\boldsymbol{\Theta}[k]) \right|. \tag{5.20}$$

for the set of DPD coefficients $\boldsymbol{\Theta} = [\, \boldsymbol{\Theta}[1] \ldots \boldsymbol{\Theta}[K] \,]$, to incorporate error weighting in both frequency and space. Here, $\boldsymbol{h}^{(q)}$ is the channel vector from branches l to direction q and $w_{q,k}$ are weights for the frequency sample k and spatial sample q. A higher $w_{q,k}$ penalizes distortion at a particular frequency k, and in a particular point in space q, more. Using (5.20), and a frequency-domain DPD implementation as in [36], Paper C shows that in-band and out-of-band distortion can be shaped separately to better yield a desired distribution of in-band and out-of-band distortion in space.

CHAPTER 6

Summary of included papers

This chapter provides a summary of the included papers.

6.1 Paper A

Björn Langborn, Christian Fager, Rui Hou, Thomas Eriksson
Improved Digital Predistortion for Concurrent Multiband Transmission Using Frequency Relocation
Published in IEEE Microwave and Wireless Technology Letters, vol. 33, no. 7, pp. 1071–1074, March 2023.

In Paper A, a DPD schematic adapting bandwise finite impulse response filters to compensate for wideband PA variations in a concurrent multi-band DPD is proposed, and experimentally validated.

Björn contributed with the main parts of the research, implementation, experimental validation and writing. All coauthors contributed significantly with internal paper review. As for technical contributions, Rui contributed with input on the DPD schematic and practical DPD considerations. Christian contributed with help in experimental work, and early-stage idea refine-

ment. Thomas contributed with implementation ideas and practical DPD considerations.

6.2 Paper B

Björn Langborn, Christian Fager, Rui Hou, and Thomas Eriksson
Concurrent multi-beam digital pre-distortion using FFT beamforming and virtual arrays
Submitted to Sensors, special issue on "Advancements in Power Amplifier Design and Linearization Techniques for Wireless Communication Systems". Available as preprint, doi: 10.20944/preprints202602.0713.v1.

In Paper B, a novel DPD scheme for concurrent multi-beam transmission in MIMO arrays is proposed. It is especially useful for large transmit arrays, and utilizes a so-called virtual array to perform DPD in a lower-dimensional space than the physical array, thus lowering DPD computational complexity. Mathematical motivation and simulation results showcase the validity and benefit of the proposed approach, showcasing manyfold reductions in DPD complexity.

Björn contributed with the main parts of the research, implementation and writing. All coauthors contributed significantly with internal paper review. As for technical contributions, Christian contributed with early stage idea refinement. Rui contributed with the initial research idea, and along the way with discussions on the mathematical foundation. Thomas contributed with help in generalization, and broadening the applicability of the method, alongside refining mathematical clarity.

6.3 Paper C

Björn Langborn, Siqi Wang, Thomas Eriksson
Distortion shaping in space and frequency using digital predistortion in MIMO transmitters
Submitted to IEEE Communications Letters.

In Paper C, a novel approach to frequency and spatial distortion shaping for nonlinear MIMO transmitters is proposed. By defining a new metric,

and adaptation of a frequency-selective DPD scheme jointly optimized across branches, it is shown that in-band and out-of-band distortion can be shaped in how they radiate spatially.

Björn contributed with the main parts of the research, implementation and writing. All coauthors contributed significantly with internal paper review. As for technical contributions, Siqi contributed with advice on system modeling and the simulation setup. Thomas contributed with early stage idea refinement and implementation.

CHAPTER 7

Concluding Remarks and Future Work

This thesis concludes that there is great potential of complexity reduction, often with little to no performance loss, in sparse multi-user settings. In the frequency-sparse, concurrent multi-user setting, state-of-the-art DPD research has focused on trading a single high sampling rate SISO-DPD to either a medium-rate SISO-DPD with pre-processing, or to a low-rate MIMO-DPD. The largest reduction in complexity, and best performance, can be obtained when carrier frequencies are equally spaced. Considering instead the spatially sparse, concurrent multi-beam setting, some conclusions from the frequency-sparse setting are transferable, whilst some are not. Linearization sparseness in the concurrent multi-beam setting can be achieved by exchanging many SISO-DPDs, to either a medium amount of SISO-DPDs, or to a single MIMO-DPD. Maximal complexity reduction is obtained when users are evenly spaced on the spatial FFT-beamgrid, seen in Paper B.

In contrast between the settings, downsampling in the time-domain, as done in the frequency relocation approach of Paper A, results in decreased ability to model short time-scale memory effects. In the spatial downsampling approach of Paper B, no time-frequency information is lost, and as such no ability to model memory effects. Additionally, the computational sparsity of the system

is increased by neglecting higher order IM terms. In practice this can most often be justified by observing that high order IM terms can be comparable in terms of power to the noise floor, although for highly nonlinear PAs this might hold less true.

There are likely many avenues for future work on these topics, in particular for utilizing spatial sparseness which so far has seen less research. A few research questions that could be of interest for future work include:

- How can practical PA variations in a MIMO array best be handled when not doing per-PA DPD? Can PA design and knowledge be utilized to not deteriorate performance in such systems, or perhaps even improve it?
- Can the formulation for finding a virtual DPD domain be rephrased, such that an analytically closed expression can be found - relating the beam indices and domain size in the virtual domain to those in the original domain?
- Can the dimensionality reduction of the virtual domain, of Paper B, be better performed using e.g. non-uniform downsampling in the antenna space?
- What avenues for complexity reduction are available whilst performing spatial and frequency distortion shaping, by the metric or approach of Paper C?

References

[1] Ericsson, "Ericsson Mobility Report June 2025," Tech. Rep., Jun. 2025.

[2] A. Katz, J. Wood, and D. Chokola, "The Evolution of PA Linearization: From Classic Feedforward and Feedback Through Analog and Digital Predistortion," *IEEE Microwave Magazine*, vol. 17, no. 2, pp. 32–40, Feb. 2016, ISSN: 15273342.

[3] K. Yamanaka, S. Shinjo, Y. Komatsuzaki, S. Sakata, K. Nakatani, and Y. Yamaguchi, "Overview and Prospects of High Power Amplifier Technology Trend for 5G and beyond 5G Base Stations," *IEICE Transactions on Electronics*, vol. E104.C, no. 10, p. 2021MMI0008, Oct. 2021, ISSN: 0916-8524.

[4] A. Sufyan, K. B. Khan, O. A. Khashan, T. Mir, and U. Mir, "From 5G to beyond 5G: A Comprehensive Survey of Wireless Network Evolution, Challenges, and Promising Technologies," *Electronics*, vol. 12, no. 10, p. 2200, May 2023, ISSN: 2079-9292.

[5] M. D. Pozar, *Microwave and RF Design of Wireless Systems*, 1st ed. John Wiley & Sons, 2001.

[6] P.-S. Kildal, *Foundations of antenna engineering: a unified approach for line-of-sight and multipath.* Artech House, 2015.

[7] N. K. Subramani, "Physics-based TCAD device simulations and measurements of GaN HEMT technology for RF power amplifier applications," Ph.D. dissertation, Université de Limoges, 2017.

[8] C. Crespo-Cadenas, M. Jose Madero-Ayora, and J. A. Becerra, *A Volterra Approach to Digital Predistortion: Sparse Identification and Estimation*, 1st ed. Wiley-IEEE Press, 2025.

[9] H. Fan, Y. Ding, G. Goussetis, and M. J. Canavate Sanchez, "Antenna Array Driven by Non-Isolated Power Amplifiers for MIMO Applications," in *2019 49th European Microwave Conference (EuMC)*, IEEE, Oct. 2019, pp. 468–471, ISBN: 978-2-87487-055-2.

[10] F. M. Barradas, P. M. Tomé, J. M. Gomes, T. R. Cunha, P. M. Cabral, and J. C. Pedro, "Power, Linearity, and Efficiency Prediction for MIMO Arrays With Antenna Coupling," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 12, pp. 5284–5297, Dec. 2017, ISSN: 0018-9480.

[11] M. O'Droma and Lei Yiming, "A New Bessel-Fourier Memoryless Nonlinear Power Amplifier Behavioral Model," *IEEE Microwave and Wireless Components Letters*, vol. 23, no. 1, pp. 25–27, Jan. 2013, ISSN: 1531-1309.

[12] X. Hu, Z. Liu, X. Yu, *et al.*, "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3923–3937, Aug. 2022, ISSN: 2162-237X.

[13] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proceedings of the IEEE*, vol. 69, no. 12, pp. 1557–1573, 1981, ISSN: 0018-9219.

[14] J. C. Pedro and S. A. Maas, *A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches*, Apr. 2005.

[15] A. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, 1981.

[16] A. Zhu and T. Brazil, "Behavioral modeling of RF power amplifiers based on pruned volterra series," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 12, pp. 563–565, Dec. 2004, ISSN: 1531-1309.

[17] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic Deviation Reduction-Based Volterra Behavioral Modeling of RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006, ISSN: 0018-9480.

[18] B. Fehri and S. Boumaiza, "Baseband Equivalent Volterra Series for Behavioral Modeling and Digital Predistortion of Power Amplifiers Driven With Wideband Carrier Aggregated Signals," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2594–2603, Nov. 2014, ISSN: 0018-9480.

[19] O. T. Demir and E. Bjornson, "The Bussgang Decomposition of Nonlinear Systems: Basic Theory and MIMO Extensions [Lecture Notes]," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 131–136, Jan. 2021, ISSN: 1053-5888.

[20] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 5, pp. 285–291, May 1976, ISSN: 0098-4094.

[21] Taijun Liu, S. Boumaiza, and F. Ghannouchi, "Deembedding static nonlinearities and accurately identifying and modeling memory effects in wide-band RF transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 11, pp. 3578–3587, Nov. 2005, ISSN: 0018-9480.

[22] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 3, pp. 980–987, Mar. 2017, ISSN: 00189480.

[23] P. L. Gilabert, A. Cesari, G. Montoro, E. Bertran, and J.-M. Dilhac, "Multi-Lookup Table FPGA Implementation of an Adaptive Digital Predistorter for Linearizing RF Power Amplifiers With Memory Effects," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 2, pp. 372–384, 2008, ISSN: 0018-9480.

[24] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 737–744, Feb. 2015, ISSN: 00189480.

[25] W. Cao and A. Zhu, "A Modified Decomposed Vector Rotation-Based Behavioral Model With Efficient Hardware Implementation for Digital Predistortion of RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 7, pp. 2443–2452, Jul. 2017, ISSN: 0018-9480.

[26] L. Ding, G. Zhou, D. Morgan, *et al.*, "A Robust Digital Baseband Predistorter Constructed Using Memory Polynomials," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 159–165, Jan. 2004, ISSN: 0090-6778.

[27] A. E. Abdelrahman, O. Hammi, A. K. Kwan, A. Zerguine, and F. M. Ghannouchi, "A Novel Weighted Memory Polynomial for Behavioral Modeling and Digital Predistortion of Nonlinear Wireless Transmitters," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 1745–1753, Mar. 2016, ISSN: 0278-0046.

[28] O. Hammi, F. Ghannouchi, and B. Vassilakis, "A Compact Envelope-Memory Polynomial for RF Transmitters Modeling With Application to Baseband and RF-Digital Predistortion," *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 5, pp. 359–361, May 2008, ISSN: 1531-1309.

[29] W. Chen, S. Zhang, Y. Liu, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Efficient Pruning Technique of Memory Polynomial Models Suitable for PA Behavioral Modeling and Digital Predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 10, pp. 2290–2299, Oct. 2014, ISSN: 0018-9480.

[30] Q. Zhang, Y. Liu, J. Zhou, S. Jin, W. Chen, and S. Zhang, "A Band-Divided Memory Polynomial for Wideband Digital Predistortion With Limited Bandwidth Feedback," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 62, no. 10, pp. 922–926, Oct. 2015, ISSN: 1549-7747.

[31] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006, ISSN: 1053587X.

[32] W. Cao, S. Wang, P. N. Landin, C. Fager, and T. Eriksson, "Complexity Optimized Digital Predistortion Model of RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 3, pp. 1490–1499, Mar. 2022, ISSN: 0018-9480.

[33] S. Wang, M. A. Hussein, O. Venard, and G. Baudoin, "A Novel Algorithm for Determining the Structure of Digital Predistortion Models," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7326–7340, Aug. 2018, ISSN: 0018-9545.

[34] J. A. Becerra, M. J. Madero-Ayora, R. G. Noguer, and C. Crespo-Cadenas, "On the Optimum Number of Coefficients of Sparse Digital Predistorters: A Bayesian Approach," *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 12, pp. 1117–1120, Dec. 2020, ISSN: 1531-1309.

[35] A. Barry, W. Li, J. A. Becerra, and P. L. Gilabert, "Comparison of Feature Selection Techniques for Power Amplifier Behavioral Modeling and Digital Predistortion Linearization," *Sensors*, vol. 21, no. 17, p. 5772, Aug. 2021, ISSN: 1424-8220.

[36] A. Brihuega, L. Anttila, and M. Valkama, "Frequency-Domain Digital Predistortion for OFDM," *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 6, pp. 816–818, Jun. 2021, ISSN: 15581764.

[37] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited volterra series-based digital predistortion for wideband RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 4198–4208, 2012, ISSN: 00189480.

[38] L. Ding, F. Mujica, and Z. Yang, "Digital predistortion using direct learning with reduced bandwidth feedback," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, IEEE, Jun. 2013, pp. 1–3, ISBN: 978-1-4673-6176-7.

[39] N. Guan, N. Wu, and H. Wang, "Digital Predistortion of Wideband Power Amplifier With Single Undersampling ADC," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 11, pp. 1016–1018, Nov. 2017, ISSN: 1531-1309.

[40] N. Kelly and A. Zhu, "Direct Error-Searching SPSA-Based Model Extraction for Digital Predistortion of RF Power Amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 3, pp. 1512–1523, Mar. 2018, ISSN: 0018-9480.

[41] J. Chani-Cahuana, C. Fager, and T. Eriksson, "A new variant of the indirect learning architecture for the linearization of power amplifiers," in *2015 10th European Microwave Integrated Circuits Conference (EuMIC)*, IEEE, Sep. 2015, pp. 444–447, ISBN: 978-2-8748-7040-8.

[42] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative Learning Control for RF Power Amplifier Linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 9, pp. 2778–2789, Sep. 2016, ISSN: 0018-9480.

[43] G. Su and W. Chen, "Digital predistortion for concurrent multi-band PAs with inter-band IMD compensation," in *2016 IEEE International Workshop on Electromagnetics, iWEM 2016 - Proceeding*, Institute of Electrical and Electronics Engineers Inc., Jul. 2016, ISBN: 9781467381949.

[44] W. Cao, S. Wang, and T. Eriksson, "Concurrent Multiband Linearization for Power Amplifiers With Intermodulation Distortions," *IEEE Microwave and Wireless Components Letters*, vol. 32, no. 5, pp. 467–470, May 2022, ISSN: 15581764.

[45] S. A. Bassam, M. H. Helaoui, and F. M. Ghannouchi, "2-D digital predistortion (2-D-DPD) architecture for concurrent dual-band transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 10 PART 1, pp. 2547–2553, 2011, ISSN: 00189480.

[46] T. Eriksson and C. Fager, "Digital predistortion of concurrent multiband communication systems," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2014, pp. 3918–3922, ISBN: 978-1-4799-2893-4.

[47] M. Younes, A. Kwan, M. Rawat, and F. M. Ghannouchi, "Three-Dimensional digital predistorter for concurrent tri-band power amplifier linearization," in *IEEE MTT-S International Microwave Symposium Digest (MTT)*, IEEE, 2013, pp. 1–4, ISBN: 9781467321419.

[48] Y. Liu, C. Li, X. Quan, *et al.*, "Multiband Linearization Technique for Broadband Signal With Multiple Closely Spaced Bands," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 3, pp. 1115–1129, Mar. 2019, ISSN: 0018-9480.

[49] C. Yu, J. Xia, X. W. Zhu, and A. Zhu, "Single-Model Single-Feedback Digital Predistortion for Concurrent Multi-Band Wireless Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 7, pp. 2211–2224, Jul. 2015, ISSN: 00189480.

[50] X. Liu, W. Chen, J. Chu, F. M. Ghannouchi, and Z. Feng, "Multi-Stream Spatial Digital Predistortion for Fully-Connected Hybrid Beamforming Massive MIMO Transmitters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 7, pp. 2998–3011, Jul. 2021, ISSN: 15580806.

[51] A. Brihuega, L. Anttila, and M. Valkama, "Beam-Level Frequency-Domain Digital Predistortion for OFDM Massive MIMO Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 4, pp. 1412–1427, 2023.

[52] C. Mollen, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transceiver nonlinearities," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6663–6679, Oct. 2018, ISSN: 15582248.

[53] A. Molina, K. Rajamani, and K. Azadet, "Concurrent Dual-Band Digital Predistortion Using 2-D Lookup Tables with Bilinear Interpolation and Extrapolation: Direct Least Squares Coefficient Adaptation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 4, pp. 1381–1393, Apr. 2017, ISSN: 00189480.

[54] S. Wang, W. Cao, R. Hou, and T. Eriksson, "A Digital Predistortion for Concurrent Dual-Band Power Amplifier Linearization Based on Periodically Nonuniform Sampling Theory," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 1, pp. 466–475, Jan. 2022, ISSN: 15579670.

[55] K. Hausmair, U. Gustavsson, C. Fager, and T. Eriksson, "Modeling and Linearization of Multi-Antenna Transmitters Using Over-the-Air Measurements," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, May 2018, pp. 1–4, ISBN: 978-1-5386-4881-0.

[56] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G massive MIMO wireless transmitters based on indirect identification of power amplifier behavior with OTA tests," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 1, pp. 315–327, Jan. 2020, ISSN: 15579670.

[57] J. Fernandez, L. Anttila, K. Buisman, *et al.*, "Over-the-Air Linearization of Phased Array Transmitters Affected by Load Modulation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 72, no. 5, pp. 2366–2379, May 2025, ISSN: 1549-8328.

[58] N. Tervo, B. Khan, J. P. Aikio, *et al.*, "Combined Sidelobe Reduction and Omnidirectional Linearization of Phased Array by Using Tapered Power Amplifier Biasing and Digital Predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 9, pp. 4284–4299, Sep. 2021, ISSN: 15579670.

[59] Y. Wang and Z. Luo, "Optimized Iterative Clipping and Filtering for PAPR Reduction of OFDM Signals," *IEEE Transactions on Communications*, vol. 59, no. 1, pp. 33–37, Jan. 2011, ISSN: 0090-6778.

[60] R. N. Braithwaite, "A Combined Approach to Digital Predistortion and Crest Factor Reduction for the Linearization of an RF Power Amplifier," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 291–302, Jan. 2013, ISSN: 0018-9480.

[61] S. Wang, M. Roger, and C. Lelandais-Perrault, "Impacts of Crest Factor Reduction and Digital Predistortion on Linearity and Power Efficiency of Power Amplifiers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 3, pp. 407–411, Mar. 2019, ISSN: 1549-7747.

[62] X. Xia, X. Quan, Y. Liu, S. Shao, and Y. Tang, "A Frequency-Selective Digital Predistortion Method Based on a Generalized Indirect Learning Architecture," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2334–2348, 2022, ISSN: 1053-587X.

[63] H. Yin and A. Zhu, "Pareto-Optimal Multimetric Model Extraction for Digital Predistortion of RF Power Amplifiers for Error Spectrum Redistribution," *IEEE Transactions on Microwave Theory and Techniques*, vol. 73, no. 4, pp. 2230–2241, 2025, ISSN: 15579670.

Part II

Papers

PAPER A

Improved Digital Predistortion for Concurrent Multiband Transmission Using Frequency Relocation

Björn Langborn, Christian Fager, Rui Hou, Thomas Eriksson

Published in IEEE Microwave and Wireless Technology Letters, vol. 33, no. 7, pp. 1071–1074, March 2023

The layout has been revised.

Abstract

In concurrent multi-band settings with large carrier spacings between bands, digital pre-distorter (DPD) implementations face the issue of how to linearize a sparse signal spectrum spanned over large bandwidths at low computational complexity. Frequency relocation is one possible approach, which involves relocating baseband carriers to have reduced frequency spacing with the help of band-limiting functions, to yield a denser and smaller signal bandwidth for the DPD to linearize. In this paper, frequency relocation for transmitters with frequency varying gain and phase is explicitly considered. It is shown that band-wise pre-equalization during frequency relocation can reduce the computational complexity for a certain allowed in-band error.

1 Introduction

Modern cellular networks employ multiple frequency bands simultaneously. The low-frequency bands have favorable propagation and penetration properties for network coverage, whereas the high-frequency bands have more abundant spectrum for network capacity. Although concurrent multi-band capability is highly desirable for wireless base-station transmitters, this capability is rarely seen in practical systems due to significant technical challenge in transmitter linearization.

Digital pre-distortion (DPD) techniques are ubiquitous, and well-performing, for linearization and compensation of non-linear distortion in radio frequency power amplifiers (PAs). DPDs enable PAs to operate with high output power and efficiency, whilst still meeting stringent linearity requirements. In contrast to wideband DPDs, which focus on linearizing a wide and contiguous spectrum, a key challenge for multi-band DPD is how to handle sparse, yet wide, bandwidths without drastically increasing DPD complexity. Another challenge is how to address intra-band, and unfilterable inter-band, intermodulation (IM) that do not arise in single-band transmitters.

DPD for single-band transmitters is a well researched topic, and it serves as a foundation for multi-band DPD. However, single-band DPD solutions are

neither computationally efficient for sparse multi-band signals, nor do they explicitly consider the aforementioned types of IM products. When it comes to DPDs specifically for the multi-band setting, solutions like the 2D-DPD and 3D-DPD have been proposed [1]–[4]. The 2D-DPD solutions consider a dual-band input and solve the issue of having sparse, yet wide, bandwidth to linearize by using a DPD per band, so that each DPD can operate at a low complexity. Intermodulation is considered by having the baseband data of each band as input to each DPD. Whilst this works well for the dual-band and tri-band cases, multi-band linearization using such an approach becomes increasingly complex in terms of e.g. the number of basis functions and number of DPDs as the number of bands increase.

A more recent paper presented the concept of frequency relocation for concurrent multi-band linearization [5]. The main idea of frequency relocation is to relocate bands that are widely spread apart to lie closer to one another, since this allows the DPD to operate on data sampled at a lower rate, which reduces computational complexity. However, considering that the gain and phase response of a transmitter can vary significantly over large frequency ranges, the frequency relocated system response will exhibit the same variations but over a smaller bandwidth. This poses a challenging compensation problem for the DPD. This paper seeks to address how transmitter gain and phase variations can be taken into consideration, by introducing band-wise pre-equalization blocks in the schematic.

In the paper that follows, Section 2 describes frequency relocation in more detail. In Section 3, the proposed frequency relocation scheme using pre-equalization coefficients is presented. Lastly, Section 4 elaborates on the experimental setup and results to verify the proposed approach.

2 Frequency relocation system description

A schematic outlining the steps involved in frequency relocation can be seen in Fig. 1. Starting from the right side, the goal is to produce an output y as similar as possible to y_d, the desired output. Given that the PA has some gain variation over frequency, indicated in solid lines in the graph above the PA, and some non-linearity, then a specific u must be input to achieve y at the output. It is seen from the schematic that this u is produced by relocating the bands in u'. Furthermore, u' is obtained by a non-linear mapping from y'_d

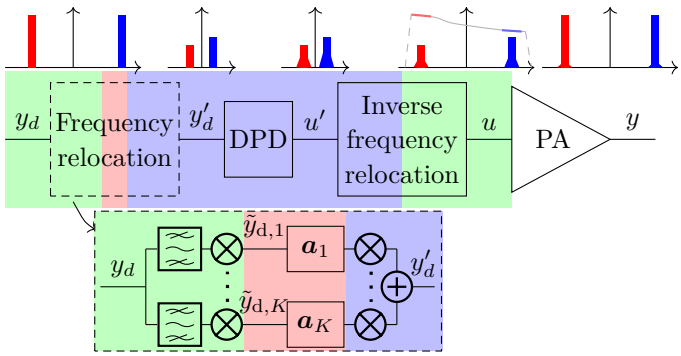


Figure 1: Simplified schematic of the frequency relocation concept, with the proposed modification elaborated below. The lines in the u spectrum indicate the gain profile of the PA. The colored regions refer to the sample rates of signals in that region. Here, red, blue and green corresponds to low, medium and high sample rates respectively. The symbols $\boldsymbol{a}_1, ...\boldsymbol{a}_K$ are vectors, and refer to FIR filter coefficients for pre-equalization.

using the DPD, and y'_d obtained by relocating the input signal y_d.

The issue with wide, yet sparse, bandwidth is here solved by relocating all bands to a lower frequency range during linearization. Additionally, IM products are accounted for by the choice of frequency spacing, as detailed in [5]. Briefly stated, if the original frequency spacing is even then the relocated frequencies are chosen to be evenly spaced as well, so that all IM end up band-centered. A bandwidth large enough for characterizing the PA's response around each band is chosen, which has been done in previous works as well as only frequency components around each carrier are typically significant for the PA output formulation [6], [7]. The separation of relocated bands is kept to have non-overlapping spectral regrowth. If the original frequency spacing is uneven, IM will end up out-of-band and one typically places bands as close as possible without getting IM partially in-band.

3 Proposed DPD schematic

A key observation to note from Fig. 1 is the fact that the PA's complex gain can vary across different frequencies. These variations typically vary smoothly over a continuous range of frequencies, and within each band the variations are

likely not very large. However, for large carrier spacings these variations can yield significantly different gain and phase responses for the different bands in a multi-band signal. These observations are exemplified in Figure 2b, and it can be understood that the optimal input u to the PA will reflect these gain variations. Furthermore, when inversely relocating the signal u to u' it is clear that the gain transition between bands in u' becomes more abrupt. So if the bands in y'_d are unaltered from their magnitudes and phases in y_d, the DPD has to compensate for the gain transition, in addition to the effects that occur with a more frequency flat response of the PA. From the time-frequency duality it can be realised that if the DPD has to compensate a sharp transition in the frequency domain, this corresponds to that the DPD requires a longer memory in time. Thus, the DPD complexity increases. In [5], no explicit consideration to this observation is made.

However, it was noted that the variations within each band are not expected to be very large. So if the abrupt changes in gain and phase between bands are handled band-wise, i.e. in a small frequency region around each carrier instead of over the whole frequency span, we expect that it can be done at a very low complexity per band. Thus, the novelty of this work is to address the sharp transitions by using band-wise pre-equalization filters to get y'_d, to make the DPD mapping from y'_d to u' easier. As detailed in the experimental results, this can both improve performance and reduce complexity compared to the original frequency relocation technique.

The proposed schematic is presented in Fig. 1. Compared to [5], FIR filters $\boldsymbol{a}_1, ...\boldsymbol{a}_K$ for each band are introduced here, that can adjust adjust the signal magnitudes and phases in y'_d. Worth noting is that if the PA gain and phase is equal in each band then the optimal choice of coefficients are $\boldsymbol{a}_i = 1$, $i = 1, .., K$ and the proposed method becomes equivalent to [5].

Determining the coefficients in $\boldsymbol{a}_1, ..., \boldsymbol{a}_K$ requires joint optimization with the DPD coefficients. The normalized mean square error (NMSE) $\epsilon_{y_d} = \Sigma_i(y_d[i] - y[i])^2/\Sigma_i y_d^2[i]$ is chosen as the performance metric to be minimized, and the minimization is done iteratively. First, an initial choice of $\boldsymbol{a}_1, ..., \boldsymbol{a}_K$ can be made e.g. based on the linear gain characteristics of the PA. Secondly, linear least squares optimization is performed for the DPD to minimize $\Sigma_i(y'_d[i] - u'[i])^2/\Sigma_i y'^2_d[i]$. Then, ϵ_{y_d} can be computed after measuring y. After this, $\boldsymbol{a}_1, ..., \boldsymbol{a}_K$ are updated using numerical optimization [8], and the iteration starts over. The goal is to find the best $\boldsymbol{a}_1, ..., \boldsymbol{a}_K$ to simplify the DPD

Table 1: ACLR and NMSE from quad-band measurement with no DPD, frequency relocation without or with a pre-equalization coefficient per band, or without frequency relocation, for different GMP orders. Scalar, band-wise, post-equalization is used for each measurement. For compactness, each ACLR value corresponds to the side of the band with most power leakage.

ACLR [dBc] **NMSE** [dB]	No DPD				DPD w/o frequency relocation				Yu et al.[5]				This work			
	2.98 GHz	3.23 GHz	3.48 GHz	3.73 GHz	2.98 GHz	3.23 GHz	3.48 GHz	3.73 GHz	2.98 GHz	3.23 GHz	3.48 GHz	3.73 GHz	2.98 GHz	3.23 GHz	3.48 GHz	3.73 GHz
P=7,M=3,					-49.1	-48.9	-49.0	-48.5	-47.8	-48.1	-46.4	-48.7	-48.6	-47.2	-47.3	-48.4
G=0	-39.1	-38.7	-36.9	-37.6	**-41.9**	**-41.6**	**-42.3**	**-41.5**	**-38.2**	**-40.1**	**-37.4**	**-36.6**	**-39.5**	**-39.0**	**-39.2**	**-39.2**
P=7,M=3,	**-31.2**	**-30.7**	**-29.1**	**-29.8**	-50.5	-49.3	-49.6	-49.9	-48.9	-49.0	-49.6	-49.3	-49.1	-49.4	-48.6	-49.4
G=1					**-43.7**	**-42.0**	**-43.0**	**-43.1**	**-36.1**	**-39.3**	**-37.9**	**-41.1**	**-41.6**	**-40.7**	**-40.6**	**-41.9**
P=7,M=5,					-50.6	-49.3	-49.7	-49.9	-50.4	-49.5	-50.2	-49.9	-49.9	-49.3	-48.8	-49.9
G=1					**-43.7**	**-42.1**	**-42.9**	**-43.2**	**-42.6**	**-39.9**	**-37.9**	**-42.6**	**-42.1**	**-41.5**	**40.8**	**-42.2**

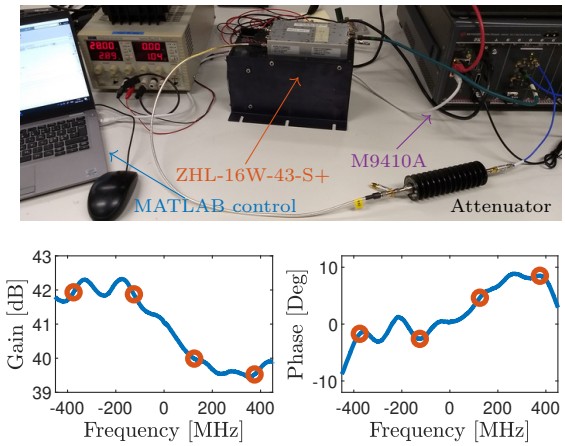
mapping from y'_d to u', resulting in a lower ϵ_{y_d} for a given DPD complexity.

Whilst the proposed scheme is a special case of a general cascade DPD structure, previous work such as [9]–[12] are not ideally suited to, with low complexity, address the aggravation of memory effects and gain transitions caused by frequency relocation. Thereby, this work is of use to reduce complexity for frequency relocation DPDs with sparse multi-band signals.

4 Experimental setup and results

To demonstrate the proposed schematic, a Keysight M9410A VXT PXI Vector Transceiver has been used to transmit signals through a MiniCircuits ZHL-16W-43-S+ power amplifier, as seen in Figure 2a. The Peak-to-Average Power Ratio of y_d was reduced to 9 dB using iterative clipping-and-filtering. Iterative Learning Control [13] was adopted for finding an optimal input signal u to linearize the PA, after which Generalized Memory Polynomial (GMP) [6] models of varying orders were adapted to map y'_d to u'. The GMP models are characterized by their non-linearity order P, memory depth M and cross-memory depth G. Using all non-linearity terms, the coefficient complexity of the DPD can be calculated as

$$C_{\text{DPD}} = P(M+1)(2G+1)\,, \tag{A.1}$$



(b)

Figure 2: Experimental setup for the DPD measurements in 2a, and PA gain characterization in 2b. For 2b, the linear gain and phase (magnitude and phase of S_{21}) of the PA at center carrier 3.35 GHz, is viewed in the baseband domain. The dots highlight the band placement for the results in Table 1.

whilst the total coefficient complexity becomes

$$C_{\text{tot}} = P(M+1)(2G+1) + KM_{\text{p}}\,, \tag{A.2}$$

where M_{p} is the pre-equalization filter length in each of the K bands. For the following results, $M_{\text{p}} = 1$ was used.

A complex baseband signal y_d, containing four frequency spaced 10 MHz-bands centered around -375, -125, 125 and 375 MHz was sent from Matlab to the M9410A, upconverted with a carrier of 3.35 GHz and transmitted. Relo-

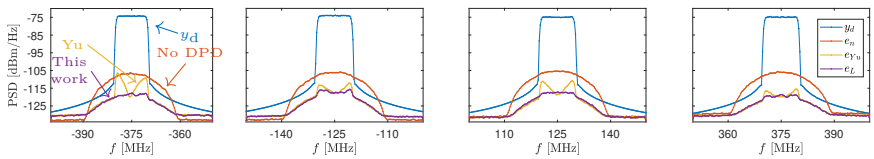


Figure 3: Power spectral density (PSD) of the desired quad-band signal y_d (in blue), and the errors $e_{\text{Yu}} = y_{\text{Yu}} - y_d$, $e_{\text{L}} = y_{\text{L}} - y_d$, $e_n = y_n - y_d$ in each band, seen in yellow (Yu et. al method [5]), purple (this work) and red (no DPD) respectively. The GMP order is $P = 7$, $M = 3$, $G = 1$.

cated frequencies were chosen as -75, -25, 25 and 75 MHz. The low, medium and high sampling rates (corresponding to red, blue and green regions in Figure 1) were chosen as 50 MHz, 200 MHz and 1200 MHz.

In Figure 3, a desired quad-band output signal (in blue) and the errors $e = y - y_d$ for the cases of no DPD (in red), Yu et. al.'s method [5] (in yellow) and the proposed method (in purple) are shown. The corresponding NMSE, and largest adjacent channel leakage ratio (ACLR) for each band, are found in Table 1. Firstly, it is seen that the addition of a pre-equalization coefficient per band improves linearization performance when the GMP complexity remains the same, comparing this work to [5]. Secondly, it can be seen that not even for a GMP of order P=7, M=5, G=1 can the method of [5] reach an NMSE below -39 dB for all bands, whilst this is achieved already at P=7, M=3, G=0 for this work. Using (A.1), the case P=7, M=5, G=1 results in 126 DPD coefficients. For the proposed method, using (A.2) the case P=7, M=3, G=0 results in a total of 32 coefficients. Thirdly, it is seen that linearization without frequency relocation performs best, though it requires 6 times higher sample rate for the DPD data, which may be practically unfeasible.

Finally, the results of a dual-band measurement varying M_p can be seen in Table 2. Here, the bandwidth of each signal is 50 MHz and the bands are centered around -220 and 220 MHz respectively, with a carrier frequency of 3.35 GHz. Relocated frequencies were -125 and 125 MHz, and the sample rates chosen as 250, 520 and 720 MHz respectively. Relocated frequencies were -125 and 125 MHz, and the low, medium and high sample rates chosen as 250, 520 and 720 MHz respectively.

Firstly, it can be seen from Table 2 that in terms of NMSE, increasing M_p can move some of complexity from the DPD to the FIR-filters, comparing especially the case Yu et al. $M = 3$ against the case $M = 2$, $M_p = 3$ for the proposed approach. That can be advantageous, since the FIR-filter computations are performed at a lower sample rate and requires no additional cost for generating basis functions [14]. Secondly, it can be seen that cases occur where it is more computationally efficient to increase the FIR-depth than to increase the DPD memory. Looking at the case $M = 3, M_p = 3$, it corresponds to a coefficient complexity of 34 according to (A.2). In comparison, $M = 6$ corresponds to a complexity of 49 according to (A.1).

Table 2: ACLR and NMSE from dual-band measurement using no DPD, and frequency relocation without or with pre-equalization with M_p coefficients per band. The GMP memory M varies, but with fixed $P = 7$, $G = 0$. Scalar, band-wise, post-equalization is used for each measurement. For compactness, each ACLR value corresponds to the side of the band with most power leakage.

ACLR [dBc] **NMSE** [dB]	No DPD	Yu et al. [5]			This work		
		M=2	M=3	M=6	M=2, M_p=1	M=2, M_p=3	M=3, M_p=3
3.13	-36.2	-45.8	-46.3	-46.4	-46.8	-46.8	-47.7
GHz	**-27.9**	**-30.4**	**-38.0**	**-39.4**	**-37.2**	**-41.1**	**-42.2**
3.57	-34.8	-45.7	-46.4	-46.5	-44.7	-44.8	-45.9
GHz	**-27.3**	**-34.4**	**-40.5**	**-40.6**	**-37.7**	**-38.0**	**-39.3**

5 Conclusion

Using band-wise pre-equalization in a frequency relocating DPD linearization scheme, the concurrent multi-band linearization performance can be improved at a lower coefficient complexity. The method improves upon previous work [5] when the PA gain and phase vary for different carriers.

References

[1] Y. Liu, W. Chen, J. Zhou, B. H. Zhou, and F. M. Ghannouchi, "Digital predistortion for concurrent dual-band transmitters using 2-D modified memory polynomials," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 281–290, 2013, ISSN: 00189480.

[2] S. A. Bassam, M. H. Helaoui, and F. M. Ghannouchi, "2-D digital predistortion (2-D-DPD) architecture for concurrent dual-band transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 10 PART 1, pp. 2547–2553, 2011, ISSN: 00189480.

[3] M. Younes, A. Kwan, M. Rawat, and F. M. Ghannouchi, "Three-Dimensional digital predistorter for concurrent tri-band power amplifier linearization," in *IEEE MTT-S International Microwave Symposium Digest (MTT)*, IEEE, 2013, pp. 1–4, ISBN: 9781467321419.

[4] E. Zenteno and D. Ronnow, "MIMO Subband Volterra Digital Predistortion for Concurrent Aggregated Carrier Communications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 3, pp. 967–979, Mar. 2017, ISSN: 00189480.

[5] C. Yu, J. Xia, X. W. Zhu, and A. Zhu, "Single-Model Single-Feedback Digital Predistortion for Concurrent Multi-Band Wireless Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 7, pp. 2211–2224, Jul. 2015, ISSN: 00189480.

[6] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006, ISSN: 1053587X.

[7] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-Band Complexity-Reduced Generalized-Memory-Polynomial Power-Amplifier Digital Predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 6, pp. 1763–1774, Jun. 2016, ISSN: 00189480.

[8] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, Jan. 1998, ISSN: 1052-6234.

[9] S. Wang, M. Abi Hussein, O. Venard, and G. Baudoin, "Optimal Sizing of Two-Stage Cascaded Sparse Memory Polynomial Model for High Power Amplifiers Linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 9, pp. 3958–3965, Sep. 2018, ISSN: 0018-9480.

[10] M. Rawat, F. M. Ghannouchi, and K. Rawat, "Three-Layered Biased Memory Polynomial for Dynamic Modeling and Predistortion of Transmitters With Memory," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 3, pp. 768–777, Mar. 2013, ISSN: 1549-8328.

[11] F. M. Barradas, L. C. Nunes, J. C. Pedro, T. R. Cunha, P. M. Lavrador, and P. M. Cabral, "Accurate Linearization with Low-Complexity Models Using Cascaded Digital Predistortion Systems," *IEEE Microwave Magazine*, vol. 16, no. 1, pp. 94–103, Feb. 2015, ISSN: 1527-3342.

[12] Hyunchul Ku, M. Mckinley, and J. Kenney, "Extraction of accurate behavioral models for power amplifiers with memory effects using two-tone measurements," in *2002 IEEE MTT-S International Microwave Symposium Digest (Cat. No.02CH37278)*, IEEE, pp. 139–142, ISBN: 0-7803-7239-5.

[13] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative Learning Control for RF Power Amplifier Linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 9, pp. 2778–2789, Sep. 2016, ISSN: 0018-9480.

[14] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A Comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010, ISSN: 00189480.

PAPER B

Concurrent multi-beam digital pre-distortion using FFT beamforming and virtual arrays

Björn Langborn, Christian Fager, Rui Hou, and Thomas Eriksson

Submitted to Sensors, special issue on "Advancements in Power Amplifier Design and Linearization Techniques for Wireless Communication Systems".
Available as preprint, doi: 10.20944/preprints202602.0713.v1

The layout has been revised.

Abstract

A digital predistortion (DPD) scheme for concurrent multi-beam transmission in fully digital multiple-input, multiple-output (MIMO) systems, using fast Fourier transform (FFT) beamforming and so-called virtual array processing, is proposed. In a MIMO array with nonlinear power amplifiers (PAs), transmitting multiple beams concurrently yields intermodulation products that end up in both user and non-user directions. In the setting with few users in a large array, the array dimension will typically be much larger than the number of generated intermodulation products. At the same time, linearization per-PA is excessively costly for large arrays. This work shows that is instead possible to linearize the system by producing predistorted user beams, and non-user intermodulation products, through DPD processing in a virtual array, of a much smaller dimension than the physical array. Theoretical derivations and simulation examples show how this approach can lead to manyfold reductions in DPD complexity.

1 Introduction

Multi-antenna transmitters offer the possibility to spatially direct multiple beams[1] concurrently, which is of interest for many applications. For communication, it allows base stations to serve multiple users in different locations concurrently. For localization and sensing, utilizing multi-beam schemes can improve localization accuracy[1]. To utilize these possibilities fully in future radio systems, the number of transmit branches is expected to increase dramatically. Simultaneously, operating PAs with as high efficiency as possible always comes with a compromise in their linearity. However, generating distortion due to non-linear operation is problematic for multiple reasons. Firstly, distortion will be emitted in-band in the directions of the users, deteriorating communication link quality. Secondly, distortion will be generated both

[1]The concept of a beam in this paper will sometimes be used interchangeably with a precoded data stream, generalizing statements from line-of-sight to other channel conditions.

in-band and out-of-band into non-user directions and can interfere with other systems. As such, non-linear distortion must be considered to fulfill transmission requirements. To this end, DPD is widely adopted.

To linearize multi-antenna transmitters, numerous schemes have been proposed, as detailed in Section 1.1, in which this paper will focus on fully digital beamforming. Some works have been either concerned with cancelling distortion specifically in user directions [2], [3], or cancelling distortion going in all directions from the transmitter when one main beam is considered [4]. Recent, state-of-the-art research has also considered distortion cancellation in different beam directions whilst concurrently transmitting multiple beams, so-called beam domain DPD (BD-DPD) [5], [6].

BD-DPD for multi-beam linearization is similar to multidimensional DPDs from the concurrent multi-band setting, such as 2D and 3D-DPD [7]–[10]. These linearization approaches take multiple user data streams as input to a DPD, and combines the streams to generate a set of multi-dimensional basis functions. As such they suffer the drawback of a drastically increasing number of multi-dimensional basis functions as the number of user streams increase. One alternative approach to this in the multi-band setting is frequency relocation [11], which involves relocating bands to lie closer in frequency, and generating all the necessary basis functions for linearization through a single nonlinearity before inverse frequency relocation. No similar technique to this exists in the multi-beam setting, to the authors' best knowledge.

1.1 Previous work

Handling distortion emitted from multi-input, multi-output (MIMO) transmitters can be done differently depending on system requirements, available resources and number of users to be served simultaneously. Previous work within two approaches for considering the emitted distortion will be briefly outlined here.

A first approach to handling nonlinear distortion in MIMO transmitter systems, done in e.g. [12], is to make distortion nulls in the beamforming pattern in specific directions. However, in [12] a single-user scenario is studied, referring to [13] for the observation that in a setting with either many beams and/or multiple channel taps, the distortion tends to be quite isotropically transmitted. This glosses over the case when e.g. there are multiple, but few, beams in relation to the array size. Additionally, a nullforming approach can

prevent distortion from being transmitted in specific directions, but it will not be able to prevent distortion from being generated at the PA outputs and subsequently transmitted in many directions. For that, distortion compensation is needed.

A second approach to handling nonlinear distortion is thereby through linearization. In a fully digital beamforming MIMO system, the most straightforward approach to linearization is adopting one DPD per PA in the transmitter chain. However, whilst likely to work well for linearization it is quite costly, both economically and in terms of energy consumption as well as in complexity, as the number of transmit branches increases. To address this, researchers have suggested alternative solutions.

For a MIMO transmitter with a single beam, so-called full-angle DPD [4] has been proposed. Here, instead of linearizing L PAs with L equal-size DPDs, it is proposed to use a common DPD for all branches and to perform only minor tuning in the branches to consider branch-wise variations. It is shown to perform well, and to be able to effectively cancel distortion from the transmitter. Worth noting in this approach is that it still requires tuning for all the branches but one, that is used as reference. From their results it can be seen that the tuning boxes are not insignificant in size, as their common DPD had 21 coefficients whilst the tuning boxes each had 3-10 coefficients. Lastly, the DPD scheme considers only a single beam, whilst it in practice is of great interest to transmit multiple beams concurrently.

Several works addressing the topic of multi-beam linearization are available. In [3], the objective is set to cancel inter-user distortion arising due to concurrent multi-beam operation, which is verified to work well for two users. Similarly, a technique for cancelling inter-user distortion is proposed and verified for three users in [14]. However, none of these works consider the distortion in non-user directions.

In [5] BD-DPD was adopted to perform linearization prior to the beamformer, using the user data streams rather than the precoded data streams going to each PA. With this approach, the DPD complexity scales in relation to the number of users K and user stream intermodulation products, rather than the number of antennas L. For $K \ll L$, this can yield large complexity savings. As aforementioned though, in similarity to 2D- and 3D-DPD for multiband signals, the number of intermodulation products scales drastically with larger K. Furthermore, it results in multi-dimensional bases, e.g.

basis functions of the form $x_1 x_2 x_3^*$. This is important to note, as for how the DPD scheme can be adapted to a look-up table (LUT) implementation, which is often made in practice to reduce arithmetic computations at the expense of memory. Adaptation of multi-dimensional basis functions to LUTs requires multi-dimensional tables, which both increases memory use but also computational cost for interpolation and extrapolation operations. This is exemplified in [15] for concurrent dual-band transmission. Lastly, BD-DPD assumes equal PAs across branches. In the presence of PA variations, the performance of BD-DPD is expected to deteriorate.

1.2 Key contributions and paper organization

In this paper, a novel approach to DPD processing using a so-called virtual array is proposed, for concurrent multi-beam linearization. In essence, the virtual array processing involves passing user streams through a low-dimensional virtual transmit array structure, and a virtual channel, whose operation is dependent on that of the actual transmit array. In the actual transmit array, L branches with nonlinear PAs will generate intermodulation (IM) distortion that will be emitted over-the-air (OTA) in both user and non-user directions when the signals are passed through antennae. In the virtual array instead, L_v branches with nonlinear DPDs aim to produce IM terms to cancel out those generated in the real array, though with much fewer branches, $L_\mathrm{v} \ll L$. It has similarities to how previous work performed dimensionality reduction for the multi-band setting, utilizing frequency domain sparseness, via frequency relocation [11] or non-uniform downsampling [16]. However, this work performs dimensionality reduction for the multi-beam setting, utilizing so-called beam domain sparseness.

The specific contributions of this paper are as follows:

- A novel approach to multi-beam DPD for MIMO transmitters, performing DPD in a small virtual antenna space, is proposed and exemplified. It is shown how such an approach can lead to manyfold reductions in complexity.

- Practical considerations in implementing the proposed scheme are accounted for, through simulation and discussion, considering multi-user cases, different channels and LUT implementation.

- Simulation results verify the theoretical findings, and show the applicability of the proposed scheme, and how the DPD scales in computational complexity in relation to the number of users.

To lead up to those contributions, the paper is organized as follows. The assumed system model description is expanded upon in Section 2. Firstly, in Subsection 2.1, the model description is expanded upon in order to comprehend how distortion is spatially distributed in the beam domain during concurrent multi-beam operation. After that, in Subsection 2.2, Fast-Fourier Transform (FFT) beamforming is described. This is followed in Section 3 by an example, where the orthogonality of FFT beamforming highlights the sparsity of the beam domain after over-the-air (OTA) transmission. Section 4 goes on to describe how processing in a virtual array, smaller than the actual physical one, can produce the same linearization signals as per-PA (PP) DPD but at a lower computational cost. Lastly, numerical simulation results are presented in Section 5 to verify the proposed approach, alongside a conclusion in Section 6.

2 System modeling

The system modeling of this work focuses on a discrete time-domain, baseband description of concurrent multi-beam transmission in a fully digital multi-branch transmitter setting. Firstly by detailing how intermodulation products, arising from nonlinear PAs, can be understood in such a description, and then related to FFT-beamforming as to motivate the suitability of the proposed approach.

Concepts referred to in this work is the beam domain, and the antenna space. The beam domain is a domain encompassing signals that have propagated through a channel and, as a result of beamforming, have formed beams. In Fig. 1, this domain is indicated with the red and orange beams, that are observed over-the-air in different directions as signals $y^{(1)}, ..., y^{(Q)}$. The user streams $x_1, ..., x_K$ are contained in some of these beams, whilst others contain only IM terms. In contrast, the antenna space refers to the signals on the different transmit branches, prior to channel propagation. In Fig. 1, the antenna space encompasses the signals $u_1, ..., u_L$, $z_1, ..., z_L$ but also $f_{\mathrm{PA},1}(z_1), ..., f_{\mathrm{PA},L}(z_L)$. The spaces are as such related through multiplication with a channel or beamforming matrix. For a channel $\boldsymbol{H} \in \mathcal{C}^{Q \times L}$, a

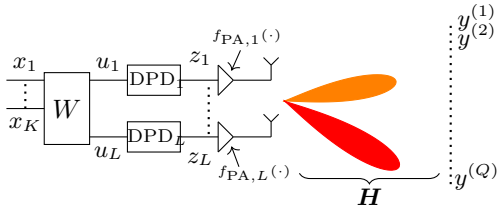


Figure 1: Concurrent multi-beam transmitter schematic of K data streams, beamformed by $\boldsymbol{W}$, passed through DPDs, transmitted through PAs on L antenna branches and observed at Q receiver positions.

signal $\boldsymbol{z} \in \mathcal{C}^{L\times N}$ in the antenna space, where N is the number of baseband data samples, the corresponding beam-domain signal is $\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{z} \in \mathcal{C}^{Q\times N}$. This relation is of particular interest when $Q = L$, and $\boldsymbol{H}$ is an orthogonal matrix, as will be later seen in the subsection on Fast-fourier transform beamforming. Additionally, whilst this paper initially expands upon the relation between the antenna space and beam domain in the context of line-of-sight (LOS) channels, it is later addressed for non-LOS conditions as well.

2.1 Concurrent transmission of multiple beams

Consider the schematic setup of Fig. 1, but start by letting the DPDs be bypassed such that $\boldsymbol{z}_l = \boldsymbol{u}_l$, $l = 1, ..., L$. In this setup, the beamformer $\boldsymbol{W}$ is formed to direct user streams $\boldsymbol{x}_1, ..., \boldsymbol{x}_K$ to directions $1, ..., Q$ in matrix form as

$$\boldsymbol{W} = \left[\, \boldsymbol{w}^{(1)} \ \ldots \ \boldsymbol{w}^{(K)} \,\right] , \tag{B.1}$$

with $||\boldsymbol{w}^{(k)}||_2 = 1$, yielding PA input signals

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_L \end{bmatrix} = [\boldsymbol{w}^{(1)}, ..., \boldsymbol{w}^{(K)}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} . \tag{B.2}$$

Alternatively, the signal z_l passed to PA l can be written as

$$z_l = w_l^{(1)} x_1 + ... + w_l^{(K)} x_K \, . \tag{B.3}$$

Now, it is well-known that an infinite Volterra series can model any non-linear function, such as PAs. For practical application however, finite truncations of

the Volterra series are instead used frequently. For qualitative analysis here, consider simply a third order polynomial baseband PA model as

$$f_{\mathrm{PA},l}(z_l) = c_{1,l} z_l + c_{3,l} z_l |z_l|^2 \,. \tag{B.4}$$

where $f_{\mathrm{PA},l}$ refers to the non-linear function of the PA in branch l, and $c_{1,l}, c_{3,l}$ are modeling coefficients. Then, inserting (B.3) into (B.4), this yields

$$\begin{aligned} f_{\mathrm{PA},l}(z_l) &= c_{1,l}(w_l^{(1)} x_1 + ... + w_l^{(K)} x_K) + \\ &\quad c_{3,l}(w_l^{(1)} x_1 + ... + w_l^{(K)} x_K)|w_l^{(1)} x_1 + ... + w_l^{(K)} x_K|^2 \\ &= c_{1,l}(w_l^{(1)} x_1 + ... + w_l^{(K)} x_K) + \\ &\quad c_{3,l}(w_l^{(1)} w_l^{(1)} w_l^{(1)*} x_1 |x_1|^2 + \\ &\qquad w_l^{(1)} w_l^{(2)} w_l^{(2)*} x_1 |x_2|^2 + \\ &\qquad w_l^{(1)} w_l^{(1)} w_l^{(2)*} x_1^2 x_2^* + ...) \,. \end{aligned} \tag{B.5}$$

First, assume that all PAs are equal, such that $c_{1,l}$ and $c_{3,l}$ are the same for each branch l. Then, let each PA output propagate through a channel vector $\boldsymbol{h}_l = [\, h_l^{(1)} \; ... \; h_l^{(Q)} \,]$ from branch l to OTA location q, as

$$y^{(q)} = [\, h_1^{(q)} \; ... \; h_L^{(q)} \,] \, [\, f_{\mathrm{PA},1}(z_1) \; ... \; f_{\mathrm{PA},L}(z_L) \,]^T \,. \tag{B.6}$$

Considering the first-order, linear terms

$$\begin{aligned} y_{\mathrm{1st}}^{(q)} &= h_1^{(q)} c_1 (w_1^{(1)} x_1 + ... + w_1^{(K)} x_K) + ... \\ &\quad + h_L^{(q)} c_1 (w_L^{(1)} x_1 + ... + w_L^{(K)} x_K) \,. \end{aligned} \tag{B.7}$$

Then it is known from beamforming theory that if x_1 is to go in direction (q), then beamforming weights can be chosen such that $w_l^{(1)} = (h_l^{(q)})^*$. This yields constructive interference $c_1 \sum h_l^{(q)} w_l^{(1)} = c_1 L$. Then, studying the third order terms in (B.5), it can be seen since $w_l^{(1)} w_l^{(1)*}, w_l^{(2)} w_l^{(2)*}$ will be real constants that $x_1 |x_1|^2$ and $x_1 |x_2|^2$ will also go in direction (q). However, the summation of the IM products last listed in (B.5), $c_3 \sum_l h_l^{(q)} w_l^{(1)} w_l^{(1)} w_l^{(2)*}$ typically has constructive interference in a direction other than (q). Generally, it can be seen that the terms of the form $\boldsymbol{w}^{(i)} \boldsymbol{w}^{(j)} \boldsymbol{w}^{(k)*}$, where $k \neq i, j$, will give rise to new beam directions of constructive interference, that are referred to as off-beams.

Table 1: Off-beam terms for static non-linearity of order $P = 3$, and with $K = 2, 3, 4$ concurrent users.

Number of users	Off-beam terms	Number of off-beam terms
$K = 2$	$x_1x_1x_2^*$, $x_2x_2x_1^*$	2
$K = 3$	$x_1x_1x_2^*$, $x_2x_2x_1^*$, $x_1x_1x_3^*$, $x_3x_3x_1^*$ $x_2x_2x_3^*$, $x_3x_3x_2^*$, $x_1x_2x_3^*$, $x_1x_3x_2^*$, $x_2x_3x_1^*$	9
$K = 4$	$x_1x_1x_2^*$, $x_2x_2x_1^*$, $x_1x_1x_3^*$, $x_3x_3x_1^*$ $x_2x_2x_3^*$, $x_3x_3x_2^*$, $x_1x_2x_3^*$, $x_1x_3x_2^*$, $x_2x_3x_1^*$, $x_1^2x_4^*$, $x_4^2x_1^*$, $x_2^2x_4^*$, $x_4^2x_2^*$, $x_3^2x_4^*$, $x_4^2x_3^*$, $x_1x_2x_4^*$, $x_1x_4x_2^*$, $x_2x_4x_1^*$, $x_1x_3x_4^*$, $x_1x_4x_3^*$, $x_3x_4x_1^*$, $x_2x_3x_4^*$, $x_2x_4x_3^*$, $x_3x_4x_2^*$	24

Two interesting observations can be made from (B.5), still assuming $c_{3,l} = c_3 \, \forall l$. Firstly, the generated off-beam terms and directions can be deterministically computed from the intended transmit directions. Off-beam terms for a static non-linearity of order $P = 3$, and with $K = 2, 3, 4$ concurrent users are listed in Table 1 to exemplify. Secondly, the non-linear products of the input streams are of a different form in the off-beams than for the in-beam. Thus, one needs to be aware of all beam directions and manage to generate the correct IM terms in the respective directions in order to linearize the system. This can readily be understood to grow more complex with more beams.

2.2 Fast-fourier transform beamforming

FFT-beamforming is one common way to computationally efficiently limit the beamforming matrix to an orthogonal set of beam directions, by first defining

$$\omega = e^{-j2\pi/L} \, . \tag{B.8}$$

Then a one-dimensional discrete Fourier transform (DFT) beamformer matrix can be formulated as

$$\boldsymbol{W}_{\text{DFT}} = \frac{1}{\sqrt{L}} \begin{bmatrix} 1 & 1 & 1 & \dots \\ 1 & \omega & \omega^2 & \dots \\ 1 & \omega^2 & \omega^4 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = \frac{1}{\sqrt{L}} \left[\, \boldsymbol{\omega}^{(b_1)} \; \boldsymbol{\omega}^{(b_2)} \; \boldsymbol{\omega}^{(b_3)} \; \dots \right] , \tag{B.9}$$

where $b_i \in [1, L]$ are beam indices, indexing the beams on the DFT-beamgrid. In implementation though, no matrix multiplication of $\boldsymbol{W}_{\mathrm{DFT}}$ with data is needed, as would otherwise be indicated by (B.2). Instead, FFT algorithms can perform this computation at lower complexity.

It is worth to note that when the main beams are limited to a DFT-beamgrid, $b_i \in [1, L]$, the off-beam directions will be restricted to the same beamgrid as well, given equal PAs. Furthermore, extension to two-dimensional beamforming is straightforward, by implementing FFT along both a horizontal and vertical direction, but omitted from the derivations for brevity. All the analysis following in this paper can then be applied to each dimension, horizontal or vertical, separately.

With the system modeling etablished, the following section details how per-PA (PP) DPD produces predistorted streams, and how intermodulation beam placement is affected by aliasing due to the antenna array being finite and discrete.

3 Per-PA DPD

In this section, the processing steps of transmitting K user streams with PP-DPD, as in Fig. 1, is detailed with a three-user example. With this example, the sparsity of the problem for larger L is highlighted, in addition to noting an effect that will be referred to as array aliasing.

3.1 Per-PA DPD example for three concurrent users

Let the antenna array be a linear, one-dimensional array in space, and restrict the beamformer to a DFT-beamgrid. Consider that there are $L = 16$ transmit branches[2], transmitting data x_1, x_2, x_3 to $K = 3$ users in different LOS directions. Being restricted to a DFT-beamgrid, the users can be located at beam indices $\{b_1, b_2, b_3\} \in [1, L]$. To exemplify, let $b_1 = 1, b_2 = 2, b_3 = 5$ and

[2]Whilst $L = 16$ is chosen for these derivations, for simpler equations, the value of L would typically be much larger in practice.

let $\boldsymbol{W}_{\mathrm{DFT}}$ be the DFT beamforming matrix to the user directions, such that

$$\begin{aligned}\boldsymbol{z} = \boldsymbol{W}\boldsymbol{x} &= \frac{1}{\sqrt{L}} \begin{bmatrix} 1 & 1 & 1 & \dots \\ 1 & \omega & \omega^2 & \dots \\ 1 & \omega^2 & \omega^4 & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ 0 \\ x_3 \\ 0 \\ \\ \vdots \end{bmatrix} \\ &= \frac{1}{\sqrt{L}} \begin{bmatrix} x_1+x_2+x_3 \\ x_1+\omega x_2+\omega^4 x_3 \\ x_1+\omega^2 x_2+\omega^8 x_3 \\ \\ \vdots \end{bmatrix} \\ &= \boldsymbol{\omega}^{(b_1)} x_1 + \boldsymbol{\omega}^{(b_2)} x_2 + \boldsymbol{\omega}^{(b_3)} x_3 \,. \end{aligned} \tag{B.10}$$

Applying the nonlinear DPDs to each of these beamformed data streams, with a third order nonlinearity $f(x) = \theta_1 x + \theta_3 x|x|^2$ as the DPD function, then

$$\underbrace{\begin{bmatrix} \theta_1(x_1+x_2+x_3)+\theta_3(x_1+x_2+x_3)|x_1+x_2+x_3|^2 \\ \theta_1(x_1+\omega x_2+\omega^4 x_3)+\theta_3(x_1+\omega x_2+\omega^4 x_3)|x_1+\omega x_2+\omega^4 x_3|^2 \\ \theta_1(x_1+\omega^2 x_2+\omega^8 x_3)+\theta_3(x_1+\omega^2 x_2+\omega^8 x_3)|x_1+\omega^2 x_2+\omega^8 x_3|^2 \\ \vdots \end{bmatrix}}_{L\times 1}, \tag{B.11}$$

is obtained. How these signals will be spatially distributed OTA can be seen by transformation to the beam domain, i.e. by simulating a channel $\boldsymbol{H}$ through an inverse Fourier transform. The user streams will then be clearly separated in the beam-domain, like the beams in Fig. 1. With the matrix $\boldsymbol{W}_{\mathrm{IDFT}} = L\boldsymbol{W}^*_{\mathrm{DFT}}$, the data streams are

$$L\boldsymbol{W}^*_{\mathrm{DFT}} f(\boldsymbol{W}\boldsymbol{x}) = \begin{bmatrix} \theta_1 x_1+\theta_3 x_1(|x_1|^2+|x_2|^2+|x_3|^2) \\ \theta_1 x_2+\theta_3 x_2(|x_1|^2+|x_2|^2+|x_3|^2) \\ \theta_3 x_2^2 x_1^* \\ \theta_3 x_1 x_3 x_2^* \\ \theta_1 x_3+\theta_3 x_3(|x_1|^2+|x_2|^2+|x_3|^2) \\ \theta_3 x_2 x_3 x_1^* \\ 0 \\ \theta_3 x_3^2 x_2^* \\ \theta_3 x_3^2 x_1^* \\ 0 \\ 0 \\ 0 \\ \theta_3 x_1^2 x_3^* \\ \theta_3 x_1 x_2 x_3^* \\ \theta_3 x_2^2 x_3^* \\ \theta_3 x_1^2 x_2^* \end{bmatrix}. \tag{B.12}$$

Each entry at a separate index here represents a signal going in a unique beam direction. Two observations are worth note. The first observations is that only 12 out of the 16 entries are non-zero. This is unsurprising, but important, as Table 1 shows that there can only be up to nine off-beam terms going in potentially different directions, in addition to the $K = 3$ user directions, for a third order nonlinearity. This is also true when L is a much larger value. Thus, for large L, there are very few low-order signal components in relation to the array size. As for the second observation, concerning where each signal component ends up on the beam grid, it is detailed in the following subsection.

3.2 Array aliasing

In (B.12), the intermodulation terms end up deterministically at a beam index by summing or subtracting the corresponding term's precoder beam indices. Exemplified, the intermodulation term $(\omega^{(b_2)}x_2)^2(\omega^{(b_1)}x_1)^* = (\omega^{(b_2)}\omega^{(b_2)}\omega^{(b_1)*})x_2^2x_1^*$ indeed produces the intermodulation term $x_2^2x_1^*$ up on beam index $b_2+b_2-b_1 = 2+2-1 = 3$. However, this can not hold for the term $(\omega^{(b_1)}x_1)^2(\omega^{(b_2)}x_2)^*$ as $2b_1 - b_2 = 0 \notin [1, L]$. Instead, this term is aliased back in to the beam grid range $[1, L]$ according to

$$\overline{b_{i_1} + b_{i_2} - b_{i_3}} = b_{i_1}+b_{i_2} - b_{i_3} - L\left\lfloor \frac{b_{i_1} + b_{i_2} - b_{i_3} - 1/2}{L} \right\rfloor, \tag{B.13}$$

for $\{b_{i_1}, b_{i_2}, b_{i_3}\} \in [1, L]$, $\{i_1, i_2, i_3\} \in [1, ..., K]$. With this in mind, it can be noted that whilst the example of Subsection 3.1 produced 12 uniquely placed data streams of main beam and intermodulation beam data, aliasing can for a different beam configuration $\{b_1, b_2, b_3\}$ end up overlapping intermodulation terms to go in the same beam directions. This will lead to distortion being concentrated in fewer beam directions.

4 Virtual array DPD

In this section, virtual array DPD (VA-DPD) processing as in Fig. 2 is proposed to reduce the complexity compared to applying a DPD per PA. As previously described in Subsection 1.2, the VA-DPD aims to produce predistorted user streams and off-beam IM terms in a low-dimensional virtual array,

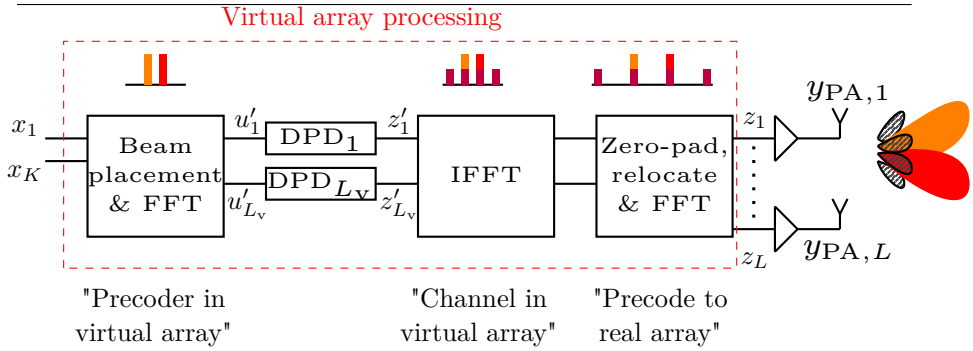


Figure 2: Schematic of DPD using virtual array processing. The bar plots exemplify the beam space representation of beamforming the data streams to beams and induced IM beams. Red and orange correspond to desired user signals, purple to the injected distortion by the DPDs to cancel out the unwanted distortion, shown in striped dark gray. Here, the number of virtual branches L_v can always be made less or equal to the number of antenna branches L.

and use these to linearize the larger, real transmit array. This is done in steps. First, user streams are precoded to the antenna space of a virtual array, by placement on a beamgrid and through an FFT. The beam placement in the virtual array depends on the precoding to be done in the real array, which will be detailed later. Second, nonlinear DPDs are applied to the data in each of the L_v virtual antenna space branches. Third, an IFFT is performed, acting as a channel and to separate user streams and off-beam IM terms in the virtual beam space. Fourth, and last, the predistorted streams that have been generated are precoded to the real array. In the following subsections, the derivations of Section 3.1 will be used as a starting point to give mathematical motivation as to how a smaller, virtual array, in place of the original array, can be used for performing DPD. This is presented first with an example comparison to the per-PA case, followed by presenting special beam configuration cases in which the complexity can be reduced even further, and finally generalizations to more general channels and higher order modeling.

4.1 Virtual array DPD example for three concurrent users

Revisiting the example of Subsection 3.1, with $\{b_1, b_2, b_3\} = \{1, 2, 5\}$ but replacing $L = 16$ with $L_\mathrm{v} = 13$ for the derivation, the computation steps are

straightforward to compute, ending with

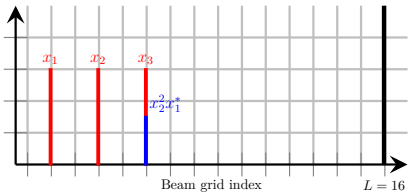
$$L_{\mathrm{v}}\boldsymbol{W}_{\mathrm{DFT}}^{*} = \begin{bmatrix} \theta_1 x_1 + \theta_3 x_1(|x_1|^2+|x_2|^2+|x_3|^2) \\ \theta_1 x_2 + \theta_3 x_2(|x_1|^2+|x_2|^2+|x_3|^2) \\ \theta_3 x_2^2 x_1^* \\ \theta_3 x_1 x_3 x_2^* \\ \theta_1 x_3 + \theta_3 x_3(|x_1|^2+|x_2|^2+|x_3|^2) \\ \theta_3 x_2 x_3 x_1^* \\ 0 \\ \theta_3 x_3^2 x_2^* \\ \theta_3 x_3^2 x_1^* \\ \theta_3 x_1^2 x_3^* \\ \theta_3 x_1 x_2 x_3^* \\ \theta_3 x_2^2 x_3^* \\ \theta_3 x_1^2 x_2^* \end{bmatrix}. \tag{B.14}$$

instead of (B.12). Comparing to Fig. 2, these computation steps have processed the data $x_1, ..., x_K$ through precoding in the virtual array domain, through DPDs and through an inverse Fourier transform. Since $L_{\mathrm{v}} < L$, fewer DPD computations have been performed whilst generating the same predistorted data streams as in (B.12). And since these data streams are separated, they can then easily be combined or precoded to an arbitrarily large physical array of size $L \geq L_{\mathrm{v}}$ by merely adding zeros and rearranging the terms.

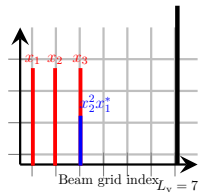
By combinatorial trials, i.e. computing (B.13) for all possible combinations of $\{b_1, b_2, b_3\}$, $L_{\mathrm{v}} = 13$ is found to be the smallest virtual domain that for any L produces all main and intermodulation streams separated as in (B.14) for $K = 3$, $P = 3$. Similarly, $L_{\mathrm{v}} = 4$ is the smallest such virtual domain for $K = 2$ users, and $L_{\mathrm{v}} = 30$ for $K = 4$. However, there are cases when the virtual domain can be made with an even smaller L_{v}, without loss of information. This will be explained further in the following subsection.

4.2 Special beam configuration cases

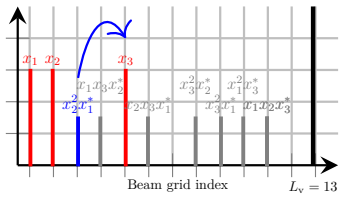
Denote $\{b'_1, b'_2, b'_3\}$ as the set of main beam indices in the virtual domain, with $\{b_1, b_2, b_3\}$ corresponding to the actual user directions. Depending on $\{b_1, b_2, b_3\}$ and L, different choices of $\{b'_1, b'_2, b'_3\}$ and L_{v} can be made to yield an optimal virtual array. More exact details on the problem of finding a virtual subspace is found in Appendix A, and is detailed statistically in the results of Section 5.3. The key phenomena explaining why this can occur is the array aliasing, as observed in Subsection 3.2, which can cause intermodulation terms to overlap. For example, evenly spaced beams in an even



(a) Beam occurrences in real array.



(b) Beam occurrences when $L_\mathrm{v} = 7$, $\{b'_1, b'_2, b'_3\} = \{1, 2, 3\}$.



(c) Beam occurrences when $L_\mathrm{v} = 13$, $\{b'_1, b'_2, b'_3\} = \{1, 2, 5\}$.

Figure 3: Beam occurrences for $\{b_1, b_2, b_3\} = \{2, 4, 6\}$ in real and virtual arrays, with placement of $x_2^2x_1^*$ highlighted.

length array produces IM terms only at even beam indices. Concretely, say $L = 16$ and $\{b_1, b_2, b_3\} = \{2, 4, 6\}$. Then, by (B.13), the beam indices for a third order nonlinearity are $[2, 4, 6, 8, 10, 14, 16]$, a total of seven unique beam directions. However, computing (B.13) with $L_v = 7$ it can be noted that with $\{b'_1, b'_2, b'_3\} = \{1, 2, 3\}$ the same IM terms will overlap with one another, but at indices $[1, 2, 3, 4, 5, 6, 7]$. This is illustrated for the main beams and one IM beam in Fig. 3a and Fig. 3b. Fig. 3c showcases how it can alternatively be obtained for $L_\mathrm{v} = 13$ by first generating all IM terms at unique indices, before moving them to the appropriate overlap index. Thus, the virtual domain size can be made smaller in special cases. The resulting virtual domain sizes in the proposed approach are as such at worst of similar size as the cardinality of unique main beam and IM beam directions.

4.3 Non-DFT beamforming

For analytical purposes, this work has focused on LOS channels, wherein the columns of the DFT precoder matrix of (B.9) describes the inverse channel for

Table 2: Fifth order IM terms for static non-linearity of order $P = 5$, and with $K = 2$ concurrent users.

Number of users	Fifth order terms	Total number of fifth order terms
$K = 2$	$x_1\|x_1\|^4$, $x_1\|x_2\|^4$, $x_1\|x_1\|^2\|x_2\|^2$ $x_2\|x_2\|^4$, $x_2\|x_1\|^4$, $x_2\|x_1\|^2\|x_2\|^2$ $x_1^3(x_2^*)^2$, $x_2^3(x_1^*)^2$ $x_1\|x_1\|^2x_2^2$, $x_2\|x_2\|^2x_1^2$ $x_1^2\|x_1\|^2x_2^*$, $x_2^2\|x_2\|^2x_1^*$	12
$K = 3$	$x_1\|x_1\|^4$, ... , $x_1\|x_2\|^2\|x_3\|^2$	60

transmission in those orthogonal LOS directions. However, more generally, if the wireless channel model can be accurately described with

$$\boldsymbol{Y} = \boldsymbol{H} f_{\mathrm{PA}}(\boldsymbol{z}) \,, \tag{B.15}$$

then the method is still applicable with some modification. Looking at the bar plots in Fig. 2, and exemplified in (B.14), it is seen that the main and IMD beam components are separated after the IFFT. Thus, the final block can be replaced with a different precoding matrix $\boldsymbol{W}$. Presume for example that some general beamforming vector $\boldsymbol{w}^{(1)}$ should be applied to spatially direct x_1, and $\boldsymbol{w}^{(2)}$ to spatially direct x_2. Seen exemplified in (B.5), the third order off-beam beamforming vectors will then be $\left(\boldsymbol{w}^{(1)}\right)^2 \left(\boldsymbol{w}^{(2)}\right)^*$ and $\left(\boldsymbol{w}^{(2)}\right)^2 \left(\boldsymbol{w}^{(1)}\right)^*$, for $x_1^2x_2^*$ and $x_2^2x_1^*$, respectively. Thus, the precoder in the final block could be computed with $\boldsymbol{W} = [\, \boldsymbol{w}^{(1)} \; \boldsymbol{w}^{(2)} \; (\boldsymbol{w}^{(1)})^2(\boldsymbol{w}^{(2)})^* \; (\boldsymbol{w}^{(2)})^2(\boldsymbol{w}^{(1)})^* \,]$ instead of through an FFT. A simple simulation case is shown in Section 5.5, with more complex cases left for future work.

4.4 Generalization to higher order terms

The analysis so far has considered third order IM beams. Extension to higher orders is very possible, by letting (B.4) include higher order terms and redoing the computations from there. This will give rise to e.g. 5th order IM terms such as $x_3^3x_1^*x_2^*$, with corresponding beam index $\overline{3b_3 - b_1 - b_2}$. As previously, some of the IM terms are always directed towards the users, and some are off-beams. Considering all these possible IM beams, the beam domain is only sparse for large L, and for VA-DPD to guarantee complete equivalence to PP-

DPD for any beam configuration, L_{v} must be increased accordingly. However, the fifth order IM terms will typically have significantly lower power than the first and third order terms. Thus, they can often be neglected in off-beam directions.

4.5 Analogy with frequency relocation

Whilst not a direct counterpart, the proposed VA-DPD approach has, as previously mentioned, similarities to e.g. frequency relocation [11]. Namely, a signal that is sparse in one of its domains is densified in that domain, prior to DPD processing. Fig. 4a shows how a multi-band signal, sparse in the frequency domain, is relocated to have frequency components closer to one another, where one when then performs DPD computations. Fig. 4b in contrast shows an original, sparse beam domain compared to its densified virtual beam space, where DPD processing is performed. However, a challenge in multi-band frequency relocation is when bands are unevenly spaced, giving rise to IM products that end up at new frequency locations. In such cases, in [11], uniform downsampling whilst maintaining the relative spacing of bands is performed. This, however, will often not densify the frequency spectrum to only contain the relevant bands and IM products. In contrast, VA-DPD can in the multi-beam setting achieve a greater degree of dimensionality reduction by careful choice of the relocated beam indices in the virtual domain.

5 Numerical setup and results

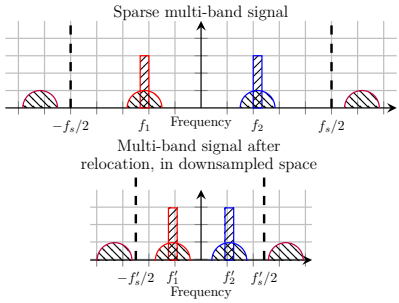
For the numerical simulations, a setup with $L = 28$ or 64, depending on the subsection, antennas in a uniform, linear array serving $K = 2$ or 3 users is considered. Each PA is modelled with a baseband Saleh model [17], equal on each PA branch, with added Gaussian noise $\boldsymbol{n}_l$ at the output as

$$f_{\mathrm{PA},l}(\boldsymbol{z}_l) = \frac{a\boldsymbol{z}_l}{1 + b|\boldsymbol{z}_l|^2} + \boldsymbol{n}_l \,, \tag{B.16}$$

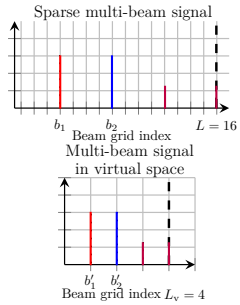
with $a = 1, b = 0.47$. The desired signal is set to

$$\boldsymbol{Y}_{\mathrm{d}} = a_d \boldsymbol{H} \boldsymbol{z} \,, \tag{B.17}$$

with $a_d = 0.95$. The noise power of $\boldsymbol{n}_l$ is chosen to give signal-to-noise ratios between 50-60 dB. $\boldsymbol{H}$ is chosen as a LOS channel, from L branches to $Q = 181$



(a) Power spectral density of a multi-band signal passed through a nonlinearity, with induced IMDs, in frequency domain.



(b) Power of a multi-beam signal passed through a nonlinearity, with induced IMDs, in beam domain.

Figure 4: Comparison of a multi-band frequency relocation approach for densifying the frequency spectrum, to the virtual array approach for densifying the beam domain.

evenly spaced observation directions in the azimuth plane. The baseband user data streams are generated through bandlimiting complex Gaussian noise to a bandwidth of 30 MHz, and normalized to a root-mean square value of 0.13, with peak-to-average power ratio about 8.3 dB. For each user stream, $N = 65536$ baseband samples are generated.

Four linearization cases are simulated per transmit configuration. Not using DPD (No DPD), using per-PA (PP) DPD, using beam-domain (BD) DPD based on [5], [6], and using virtual-array (VA) DPD.

5.1 DPD identification

With this setup, conventional iterative learning control (ILC) [18] is modified with a simple extension and used to find the optimal input signals $\boldsymbol{z}_l$ to each PA in the array. Denote

$$\boldsymbol{z}^{[i]} = \begin{bmatrix} \boldsymbol{z}_1^{[i]} \\ \vdots \\ \boldsymbol{z}_L^{[i]} \end{bmatrix}, \tag{B.18}$$

as the DPD input at iteration $[i]$. Then the DPD input at iteration $[i+1]$, $\boldsymbol{z}^{[i+1]}$, based on the OTA measured $\boldsymbol{Y}^{[i]}$ and desired signal $\boldsymbol{Y}_{\mathrm{d}}$, is found as

$$\boldsymbol{z}^{[i+1]} \leftarrow \boldsymbol{z}^{[i]} + \eta \boldsymbol{H}^{\dagger}(\boldsymbol{Y}_{\mathrm{d}} - \boldsymbol{Y}^{[i]}) \,, \tag{B.19}$$

Here, η is a step parameter for the iterations, and $\boldsymbol{H}^{\dagger} = (\boldsymbol{H}^{*}\boldsymbol{H})^{-1}\boldsymbol{H}^{*}$ is the pseudoinverse of the channel matrix $\boldsymbol{H}$. This procedure is suitable to perform offline as an initial identification, whilst online adaptation of DPD coefficients can be done with other known methods. The signals $\boldsymbol{z}_l$ are after ILC convergence shifted to the virtual domain, yielding target signals $\boldsymbol{z}_l^{'}$, as in Fig. 2. The shifting is performed by doing the VA processing steps in reverse. That is, an IFFT is performed on $\boldsymbol{z}$ along the branches, followed by that the streams of different considered beams from $\boldsymbol{z}$ are relocated to their VA-placement $\{b'_1, ..., b'_B\}$, whilst removing streams that were zero-padded in the forward path. Lastly, an FFT is done in the virtual domain to go from the virtual beam domain to the virtual antenna domain, and obtain $\boldsymbol{z}_l^{'}$. Forming a nonlinear model from u'_l to z'_l in the virtual domain through polynomial DPD models yields a predistorted virtual domain signal and basis

$$\boldsymbol{z}'_l = \boldsymbol{F}_{u'_l}\boldsymbol{\theta}_l \,, \boldsymbol{F}_{u'_l} = [\, u'_l, \, u'_l|u'_l|^2, \, ...u'_l|u'_l|^{P-1} \,] \tag{B.20}$$

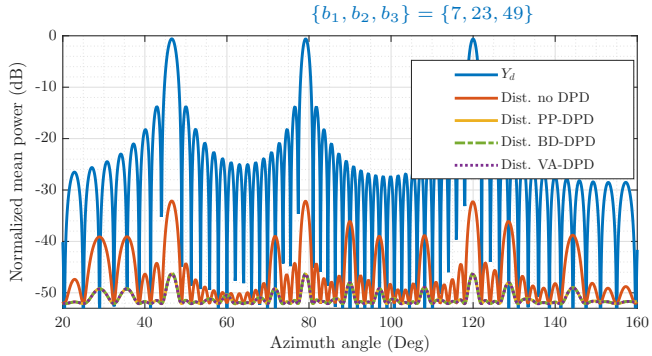
where DPD coefficients are obtained as

$$\boldsymbol{\theta}_l = \boldsymbol{F}^{\dagger}_{u'_l} z'_l \,. \tag{B.21}$$

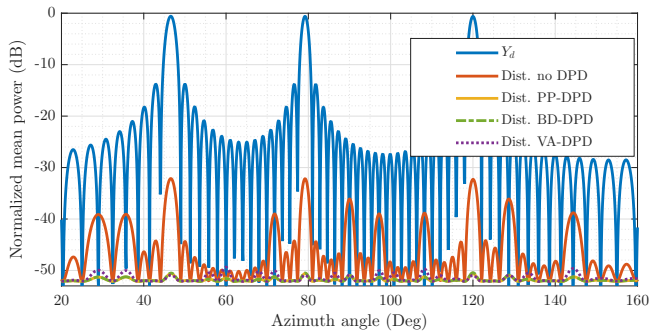
For conventional PP-DPD, equations (B.20), (B.21) are used but for the setup as in Fig. 1, with target signals $\boldsymbol{z}_l$ and inputs $\boldsymbol{u}_l$, $l = 1, ..., L$. For BD-DPD, each unique bases expressed in terms of $x_1, ..., x_K$, obtainable from the expansion of $\sum_{p=0,2,...}^{P-1} z_l|z_l|^p$, is used. This includes offbeams as in Table 1, fifth order terms as in Table 2, and main beam terms such as $x_1|x_1|^2, x_2|x_3|^2$.

5.2 Tri-beam scenario

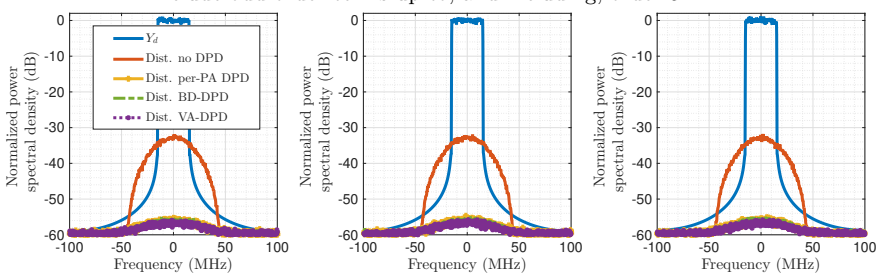
In Fig. 5, a concurrent tri-beam scenario is shown. For this scenario there are $L = 64$ transmit branches, with beams in beam directions $\{b_1, b_2, b_3\} = \{7, 23, 49\}$. In Fig. 5a and Fig. 5b, the power of the desired signal Y_{d}, over different angles, is visualized in contrast to the distortion power remaining after applying no DPD, PP-DPD, BD-DPD and the proposed VA-DPD approach with either odd order $P = 3$ or $P = 5$ DPDs. For VA-DPD, the



(a) Normalized mean power in varying spatial directions. DPDs include odd order terms up to, and including, order 3.



(b) Normalized mean power in varying spatial directions. DPDs include odd order terms up to, and including, order 5.



(c) Power spectral density in user directions. DPDs include odd order terms up to, and including, order 5.

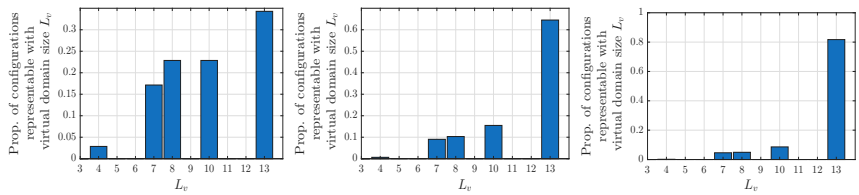
Figure 5: Normalized power of the desired signal Y_{d} shown in comparison to the distortion power remaining after no DPD, PP-DPD, VA-DPD and BD-DPD, for concurrent tri-beam transmission.

virtual array uses $L_\mathrm{v} = 13$ branches, meaning it can capture third order beam direction information without loss. Consequently, Fig. 5a, highlights how VA-DPD is equivalent in performance to PP-DPD and BD-DPD for a DPD of order $P = 3$. This is seen from that the yellow line, corresponding to PP-DPD distortion, is precisely covered by the purple and green, of VA-DPD and BD-DPD, respectively. On the other hand, compared to Fig. 5b, Fig. 5b shows that if applying fifth order DPDs whilst only considering up to third order terms in the virtual array dimensioning, the distortion power in the user directions is reduced equivalently to PP-DPD and BD-DPD, by an additional 5 dB compared to $P = 3$. However, the off-beam linearization performance is not improved correspondingly, as seen by the residual distortion e.g. around angles 28-35°.

With only $L_\mathrm{v} = 13$ branches and a first, third plus fifth order term per branch for linearization, only $13 \cdot 3 = 39$ one-dimensional DPD multiplications are needed. Alternatively, $L_\mathrm{v} = 13$ 1D-LUT look-ups are done to find the appropriate predistorted values per sample. In contrast, PP-DPD requires $64 \cdot 3$ DPD multiplications, or 64 1D-LUT look-ups. Thus, complexity is reduced by $100 \cdot |13 - 64|/64 \approx 80\%$ in VA-DPD compared to PP-DPD. Lastly, seen from Table 1 and Table 2, BD-DPD needs to account for 3 linear, 12 third order and 60 fifth order terms. For a LUT implementation, each of the twelve unique data streams, going in different spatial directions, needs a three-dimensional table look-up.

5.3 Virtual array dimensioning

The results of this section describe how small the virtual array can be dimensioned in terms of L_v, to account for all IM beams, with respect to special cases as described in Subsection 4.2. In Fig. 6 the proportion of beam configuration $\{b_1, b_2, b_3\}$ that can be represented with a virtual array of length L_v is shown for varying arrays sizes $L = 16$, 32, 64, and in Table 3 and Table 4, it is shown for various L, and $P = 3, 5$. Most noticeably, the proportion of cases requiring a larger L_v increases with L. This correlates to that in a larger beam space, the IM products end up in unique beam directions more often. The results highlight that there are cases when the beam domain is especially sparse, such that the VA-DPD can be performed at an even lower computational complexity.



(a) Array size $L = 16$. **(b)** Array size $L = 32$. **(c)** Array size $L = 64$.

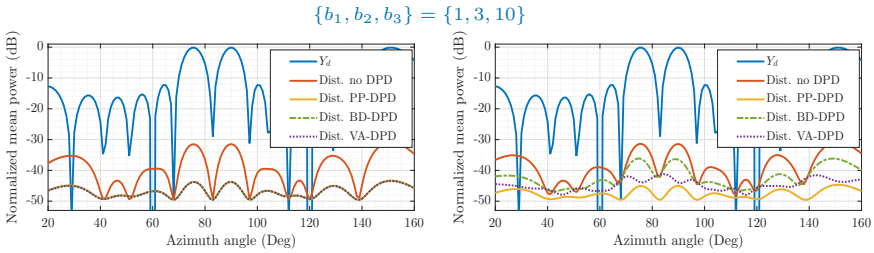
Figure 6: Proportion of main beam configuration cases in the original domain of size L that can be represented by a virtual domain of size L_v, considering IM terms of order $P = 3$.

Table 3: Percentage (%) of main beam configurations in the original domain of size L that can be represented by a virtual domain of size L_v, for IM terms of order $P = 3$.

L \ L_v	3	4	5	6	7	8	9	10	12	13
20	0	2	4	0	11	14	0	21	0	49
24	< 1	1	0	4	10	12	0	14	21	38
30	< 1	0	2	2	7	9	0	15	18	47
36	< 1	1	0	2	7	7	3	11	12	58
130	0	0	$\ll 1$	0	2	2	0	5	0	91
256	0	$\ll 1$	0	0	1	1	0	2	0	95

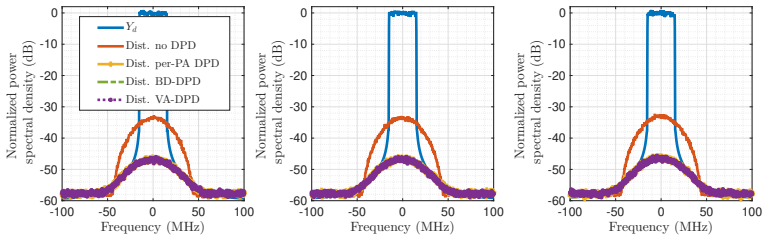
Table 4: Percentage (%) of main beam configurations in the original domain of size L that can be represented by a virtual domain of size L_v, for IM terms of order $P = 5$.

L \ L_v	4	8	11	12	16	21	22	24	26	30
32	1	4	8	8	23	10	5	21	10	10
64	< 1	1	4	4	10	7	4	15	7	47
128	$\ll 1$	< 1	2	2	5	4	2	8	4	71

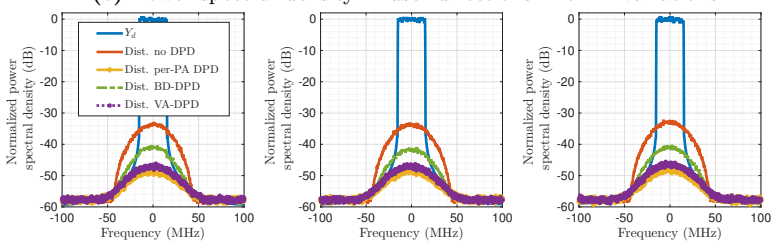


(a) Normalized mean power in varying spatial directions. No PA variations.

(b) Normalized mean power in varying spatial directions. PA variations.



(c) Power spectral density in user directions. No PA variations.



(d) Power spectral density in user directions. PA variations.

Figure 7: Normalized power of the desired signal Y_d shown in comparison to the distortion power remaining after no DPD, PP-DPD, VA-DPD and BD-DPD, for concurrent tri-beam transmission, with or without PA variations. DPDs include first and third order terms.

5.4 Power amplifier variations

It was noted earlier that BD-DPD assumes equal PAs across branches. It is this assumption that mathematically allows for a common DPD to be used

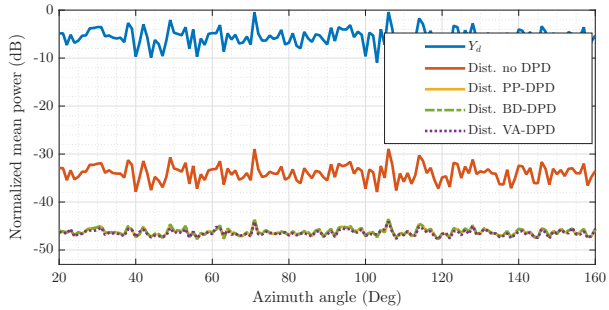
prior to the precoder in BD-DPD, without losing any ability to linearize the system [6]. This assumption is not quite as rigidly held in the proposed VA-DPD, as DPD variations across the virtual array branches can be adapted. As such, to investigate whether or not VA-DPD might handle PA variations better in comparison to BD-DPD, the PA model is changed to

$$f_{\mathrm{PA},l}(x) = c_{1,l}x + c_{3,l}x|x|^2\,, \tag{B.22}$$

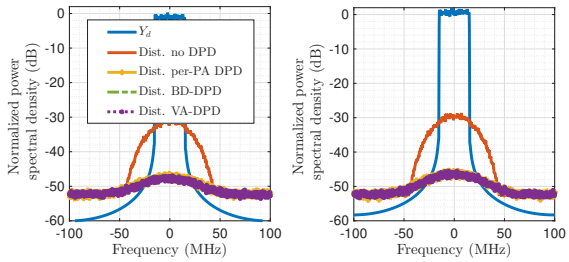
to easier represent gain and phase variations in the nonlinear PA response relative to the linear response, variations which can not be compensated by calibration. The first order coefficients are set to $c_{1,l} = 1$ for all branches, whilst $c_{3,l}$ are varied from a nominal value of -0.55 with magnitude standard deviation 0.028 and random phases from a uniform distribution between $[-20, 20]^\circ$. Fig. 7 showcases a tri-beam case with $\{b_1, b_2, b_3\} = \{1, 3, 10\}$, $L = 16$, $L_\mathrm{v} = 7$. In Fig. 7a and Fig. 7c, the PAs exhibit no variations, whilst in Fig. 7b and Fig. 7d they do. Notably, both VA-DPD and BD-DPD deteriorate in performance relative to PP-DPD with PA variations. However, VA-DPD reduces distortion more than BD-DPD in all user directions. This because whilst BD-DPD adapts one set of coefficients, prior to the precoder, to hold for linearization across all branches, VA-DPD can have some DPD coefficient variations with $l = 1, ..., L_\mathrm{v}$.

5.5 Non-FFT precoder

In Fig. 8, a two-user scenario for a random channel $\boldsymbol{H}$ of dimension $L \times Q$ consisting of complex Gaussian number distributed as $\mathcal{CN}(0, 1)$ is shown, with users in directions $[71, 106]^\circ$. The array is of size $L = 28$, and virtual array of size $L_\mathrm{v} = 4$. Pre-coding vectors are chosen as the pseudo-inverse of the respective channel vectors, i.e. $\boldsymbol{w}_1 = \boldsymbol{H}_{71}^\dagger$, $\boldsymbol{w}_2 = \boldsymbol{H}_{106}^\dagger$. First to note, all linearization methods achieve about 20 dB improvement in distortion reduction. Secondly, neither the linear power nor the residual distortion form clear beam directions, but are instead spread across the spatial range, due to the randomness of the channel. A $100 \cdot |4 - 28|/28 \approx 86\%$ reduction in DPD complexity is achieved for VA-DPD compared to PP-DPD.



(a) Normalized mean power in varying spatial directions.



(b) Power spectral density in user directions.

Figure 8: Normalized power of the desired signal Y_d shown in comparison to the distortion power remaining after no DPD, PP-DPD, VA-DPD and BD-DPD, for concurrent dual-beam transmission in random channel. $P = 7$ is used for the DPDs.

6 Conclusion

The proposed DPD schematic, using FFT beamforming and virtual array DPD processing, is shown capable of reducing DPD complexity manyfold in contrast to conventional per-PA DPD, when many transmit branches serve few users. The virtual array DPD scheme is shown operable for varying degrees of PA nonlinearities, and for different channel conditions and multi-user scenarios. In contrast to previous approaches utilizing a beam-domain framework, the proposed approach utilizes single-input, single-output DPD units that are computationally cheap to implement. Overall, the approach offers a flexible and powerful alternative to multi-user linearization for large MIMO arrays with digital beamforming.

Appendix

A Finding a virtual DPD space

Finding a virtual array representation for a given array of size L transmitting beams $b_1, ..., b_K$ involves finding an appropriate virtual array size L_v and virtual beam placement $b'_1, ..., b'_K$. In the original antenna domain, for a tri-beam case, main and intermodulation beams will end up on integer indices $c_1, ..., c_{12}$, as in

$$\begin{aligned} b_1 &= c_1 \\ b_2 &= c_2 \\ b_3 &= c_3 \\ (2b_1 - b_2) - L\left\lfloor \frac{2b_1 - b_2 - 1/2}{L} \right\rfloor &= c_4 \\ &\vdots \\ (b_2 + b_3 - b_1) - L\left\lfloor \frac{b_2 + b_3 - b_1 - 1/2}{L} \right\rfloor &= c_{12}\,, \end{aligned} \tag{B.23}$$

where $\{c_1, ..., c_{12}\} \in [1, L]$. The set $\{c_1, ..., c_{12}\}$ might not be a unique set of 12 integers. This can be exemplified with beams $\{b_1, b_2, b_3\} = \{1, 2, 4\}$, $L = 16$ where the off-beam indices are $\{16, 3, 14, 7, 16, 6, 15, 3, 5\}$ such that (B.23)

becomes

$$\begin{aligned}
b_1 &= c_1 = 1 \\
b_2 &= c_2 = 2 \\
b_3 &= c_3 = 4 \\
(2b_1 - b_2) - L\left\lfloor \frac{2b_1 - b_2 - 1/2}{L} \right\rfloor &= c_4 = c_8 = 16 \\
(2b_2 - b_1) - L\left\lfloor \frac{2b_2 - b_1 - 1/2}{L} \right\rfloor &= c_5 = c_{11} = 3 \\
&\vdots \\
(b_1 + b_3 - b_2) - L\left\lfloor \frac{b_1 + b_3 - b_2 - 1/2}{L} \right\rfloor &= c_{11} = c_5 = 3 \\
(b_2 + b_3 - b_1) - L\left\lfloor \frac{b_2 + b_3 - b_1 - 1/2}{L} \right\rfloor &= c_{12} = 5\,,
\end{aligned} \tag{B.24}$$

The objective is then to find a virtual subspace with $L_\mathrm{v}, b'_1, b'_2, b'_3$ such that

$$\begin{aligned}
b'_1 &= c'_1 \\
b'_2 &= c'_2 \\
b'_3 &= c'_3 \\
(2b'_1 - b'_2) - L_\mathrm{v}\left\lfloor \frac{2b'_1 - b'_2}{L_\mathrm{v}} \right\rfloor &= c'_4 = c'_8 \\
(2b'_2 - b'_1) - L_\mathrm{v}\left\lfloor \frac{2b'_2 - b'_1}{L_\mathrm{v}} \right\rfloor &= c'_5 = c'_{11} \\
&\vdots \\
(b'_1 + b'_3 - b'_2) - L_\mathrm{v}\left\lfloor \frac{b'_1 + b'_3 - b'_2}{L_\mathrm{v}} \right\rfloor &= c'_{11} = c'_5 \\
(b'_2 + b'_3 - b'_1) - L_\mathrm{v}\left\lfloor \frac{b'_2 + b'_3 - b'_1}{L_\mathrm{v}} \right\rfloor &= c'_{12}\,,
\end{aligned} \tag{B.25}$$

for any $\{c'_1, ..., c'_{12}\} \in [1, L_\mathrm{v}]$. This is a Diophantine system of equations, and generally not trivial to solve. However, in practice, it can be solved for

offline, and the indices obtained during runtime using a lookup table. In this work, the pseudo-algorithm of Algorithm 1 describes one approach to solving the systems of equations, for a general case with arbitrary number of beams K and branches L. Depending on K, heuristic knowledge can limit the search space Lvs and mbs in Algorithm 1. One general constraint is $L_{\mathrm{v}} \geq |\{b'_1, b'_2, ..., b'_2 + b'_3 - b'_1\}|$, i.e. the virtual domain size can never be reduced below the cardinality of the set it is supposed to represent.

Algorithm 1 Algorithm to find virtual space representation.

```
Require: mb, L // Main beam indices mb and array size L
  // Get off-beams, and cardinality of full index set
  ob = get_offbeams(mb,L)
  ab = [mb ob]
  c = length(unique(ab))
  // Try virtual subspace size Lv out of candidates Lvs
  for Lv=Lvs do
    // Try virtual main beams mbv out of candidates mbvs
    for mbv=mbvs do
      // Get off-beam indices for mbv
      obv = get_offbeams(mbv,Lv)
      abv = [mb obv]
      // Relabel ab and abv such that the first element, and
      all other elements in the vector with same value,
      is labelled 1, second element is labelled 2, etc.
      Obtain ab_relabel, abv_relabel
      if ab_relabel==abv_relabel then
        // Virtual array representation {mbv,Lv} found.
      end if
    end for
  end for
```

References

[1] A. Kakkavas, G. Seco-Granados, H. Wymeersch, M. H. C. Garcia, R. A. Stirling-Gallacher, and J. A. Nossek, "5G Downlink Multi-Beam Signal Design for LOS Positioning," in *IEEE Global Communications Conference*, IEEE, 2019, ISBN: 9781728109626.

[2] X. Liu, W. Chen, J. Chu, F. M. Ghannouchi, and Z. Feng, "Multi-Stream Spatial Digital Predistortion for Fully-Connected Hybrid Beamforming Massive MIMO Transmitters," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 7, pp. 2998–3011, Jul. 2021, ISSN: 15580806.

[3] J. Jing and C. Yu, "Multibeam Digital Predistortion for Millimeter-Wave Analog Beamforming Transmitters," *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 2, pp. 209–212, Feb. 2020, ISSN: 15581764.

[4] C. Yu, J. Jing, H. Shao, *et al.*, "Full-Angle Digital Predistortion of 5G Millimeter-Wave Massive MIMO Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 2847–2860, Jul. 2019, ISSN: 15579670.

[5] A. Brihuega, L. Anttila, and M. Valkama, "Beam-Level Frequency-Domain Digital Predistortion for OFDM Massive MIMO Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 4, pp. 1412–1427, 2023.

[6] R. N. Braithwaite, "Amplifier Nonlinearities in an Antenna Array During Spatially-Multiplexed Transmissions," in *IEEE Radio and Wireless Symposium*, IEEE, 2020, pp. 235–238.

[7] S. A. Bassam, M. H. Helaoui, and F. M. Ghannouchi, "2-D digital predistortion (2-D-DPD) architecture for concurrent dual-band transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 10 PART 1, pp. 2547–2553, 2011, ISSN: 00189480.

[8] Y. Liu, W. Chen, J. Zhou, B. H. Zhou, and F. M. Ghannouchi, "Digital predistortion for concurrent dual-band transmitters using 2-D modified memory polynomials," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 281–290, 2013, ISSN: 00189480.

[9] T. Eriksson and C. Fager, "Digital predistortion of concurrent multiband communication systems," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2014, pp. 3918–3922, ISBN: 978-1-4799-2893-4.

[10] M. Younes, A. Kwan, M. Rawat, and F. M. Ghannouchi, "Three-Dimensional digital predistorter for concurrent tri-band power amplifier linearization," in *IEEE MTT-S International Microwave Symposium Digest (MTT)*, IEEE, 2013, pp. 1–4, ISBN: 9781467321419.

[11] C. Yu, J. Xia, X. W. Zhu, and A. Zhu, "Single-Model Single-Feedback Digital Predistortion for Concurrent Multi-Band Wireless Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 7, pp. 2211–2224, Jul. 2015, ISSN: 00189480.

[12] N. Kolomvakis, M. Bavand, I. Bahceci, and U. Gustavsson, "A Distortion Nullforming Precoder in Massive MIMO Systems With Nonlinear Hardware," *IEEE Wireless Communications Letters*, vol. 11, no. 9, pp. 1775–1779, Sep. 2022, ISSN: 21622345.

[13] C. Mollen, U. Gustavsson, T. Eriksson, and E. G. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transceiver nonlinearities," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6663–6679, Oct. 2018, ISSN: 15582248.

[14] J. Jing, H. Shao, P. Yan, Z. H. Jiang, and C. Yu, "Digital predistortion of millimeter-wave multi-beam transmitters with digital beam-forming network," in *2019 IEEE MTT-S International Wireless Symposium (IWS)*, 2019, pp. 1–3.

[15] A. Molina, K. Rajamani, and K. Azadet, "Concurrent Dual-Band Digital Predistortion Using 2-D Lookup Tables with Bilinear Interpolation and Extrapolation: Direct Least Squares Coefficient Adaptation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 4, pp. 1381–1393, Apr. 2017, ISSN: 00189480.

[16] S. Wang, W. Cao, R. Hou, and T. Eriksson, "A Digital Predistortion for Concurrent Dual-Band Power Amplifier Linearization Based on Periodically Nonuniform Sampling Theory," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 1, pp. 466–475, Jan. 2022, ISSN: 15579670.

[17] A. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, 1981.

[18] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative Learning Control for RF Power Amplifier Linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 9, pp. 2778–2789, Sep. 2016, ISSN: 0018-9480.

PAPER C

Distortion shaping in space and frequency using digital predistortion in MIMO transmitters

Björn Langborn, Siqi Wang, Thomas Eriksson

Submitted to IEEE Communications Letters

The layout has been revised.

Abstract

A novel approach to multiple-input, multiple-output transmitter distortion shaping that shapes emitted distortion at different frequencies, and in different spatial directions, differently is proposed. In this approach, a new criterion of optimality using weighted over-the-air observations, and a digital predistortion (DPD) scheme particularly adapted to this criterion, are formulated. Through simulation, it is shown that emitted in-band and out-of-band distortion can be shaped spatially in the proposed approach, with up to 8 dB in-band distortion reduction in a user direction, compared to conventional per-PA DPD.

1 Introduction

To increase energy efficiency, capacity and coverage in wireless transmitters, power amplifiers (PAs) can be designed or driven to operate nonlinearly, causing distortion to be transmitted. As to not deteriorate signal quality for receiving users, and for an operator to not contaminate other operators' part of the frequency spectrum, this distortion must be considered. One approach to reduce distortion in user or potential victim directions is so-called nullforming of distortion in those directions, as done in e.g. [1][2]. This addresses the concern for deteriorating signal quality in those directions. However, emissions in other directions are not compensated. Another approach is distortion compensation via linearization, commonly performed with digital predistortion (DPD), which involves injecting compensatory distortion to reduce the overall undesired emissions and obtain a closer approximation to the desired output.

In a conventional single-input, single-output (SISO) setting, the problem of identifying a DPD is well-studied and often relatively straight-forward. The input signal is adjusted until it yields a desired, linear output signal. In a multiple-input, multiple-output (MIMO) transmitter setting, the problem is increased in dimensionality. The simplest approach to tackle this would be to treat each PA individually when linearizing, thus reducing the MIMO problem to a set of SISO optimization problems. However, this does not consider the

filtering and recombination of signals when passing through antennae and a wireless channel.

In [3], a criterion to minimize distortion going to a specific user direction is formulated in a MIMO DPD setting. However, no explicit regard to other spatial directions is taken. In [4], side-lobe reduction and omni-directional linearization is performed via a combination of PA biasing and DPD. It shows great results in shaping both side lobes, and emitted distortion. However, whilst the objective is stated to lower distortion in the main lobe, no explicit optimization criterion is stated.

In [5], [6], weights are added after performing beam-oriented DPD [7], to widen the angular region that is linearized around a user. Specifically, the nonlinear power emitted in a set of angular directions is minimized, with a constraint to keep user-oriented power constant. These works show great possibility in shaping distortion in space by utilizing the additional degrees of freedom offered in MIMO transmission compared to SISO systems. However, these works provide no precise control in forming where residual distortion is emitted, nor a way to separately control in-band (IB) and out-of-band (OOB) distortion.

Frequency-domain (FD) DPD, as in [8], has been proposed to achieve frequency selective linearization. Other works, e.g. with a time-domain approach and bandpass filtering [9], or a Pareto-optimal multimetric DPD approach [10], have also studied this problem. However, the approach of [8] lends itself very flexibly to allowing different sets of DPD coefficients to be applied on different parts of the frequency spectrum. This, in conjunction with the idea of spatial distortion shaping, lays the groundwork of this work.

2 Motivation and contribution

The motivation for performing spatial and frequency distortion shaping is manyfold. Most generally stated, it allows for antennae, channel and scenario specific information to be utilized when considering how distortion is emitted, allowing flexibility in fulfilling emission requirements. Particularly, spatial distortion shaping allows for improved performance in a set of spatial directions, at the cost of increased distortion in other directions. Combining this with a frequency selective DPD, the distortion can further be directed such that IB and OOB emissions can be allowed to different extents, in different spatial

directions. This is especially interesting in a MIMO setting. Improving the signal to noise and distortion ratio (SNDR) in a user direction can allow for higher order modulation formats to be used, enabling an increase in data rate. Simultaneously, any IB power emitted in non-user directions is undesired. As such, increasing the IB distortion, but not the total IB power, in non-user directions does not deteriorate system performance. On the other hand, it can allow for DPD complexity reductions or improvements in user directions. Lastly, if a user is known to be served with a lower order modulation format, the corresponding SNDR requirements are lower. In such a case, it can instead be beneficial to increase IB distortion for the benefit of lower OOB distortion.

Following this motivation, this work contributes as follows:

- An optimization criterion that considers the spatial importance of radiated distortion in different directions, as well providing flexibility in spatially directing IB and OOB distortion separately, is formulated.
- Details on how FD-DPD identification, particularly suited for the proposed criterion, can be performed.
- Simulation results to show the feasibility of the proposed approach.

To understand this work, the paper first details the general system modeling in section 3. Following that, section 4 discusses the criteria for DPD optimization, followed by DPD identification in section 5. Thereafter, the simulation setup and results are presented, discussed and the findings concluded.

3 System modeling

Consider a transmitter with U users with separate data streams $x_1, ..., x_U$, pre-coded through $\boldsymbol{W}$ to L transmit branches with DPDs, PAs, and then propagated through antennae and observed over-the-air (OTA) in Q directions, as in Fig. 1. Denote the precoded input signals by $v_1, ..., v_L$ and the DPD output signals by $z_1, ..., z_L$. A sample of the received baseband signal in the q:th direction OTA can be modelled as

$$y^{(q)} = \boldsymbol{h}^{(q)} f_{\mathrm{PA}}(\boldsymbol{z}) \,, \tag{C.1}$$

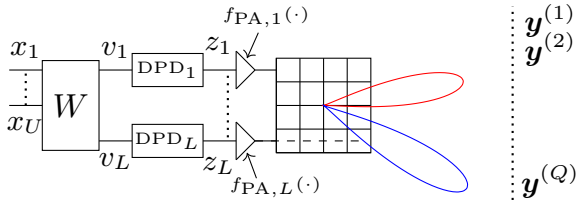


Figure 1: Transmitter with U user data streams precoded through $\boldsymbol{W}$ to L transmit branches with DPDs, PAs and propagated through antennae OTA to Q observation locations.

where it has propagated through the channel $\boldsymbol{h}^{(q)}$ and PAs $f_{\text{PA}}(\cdot)$ as

$$\boldsymbol{h}^{(q)} = \begin{bmatrix} h_1^{(q)} & \dots & h_L^{(q)} \end{bmatrix}, \tag{C.2a}$$

$$f_{\text{PA}}(\boldsymbol{z}) = \begin{bmatrix} f_{\text{PA},1}(z_1) & \dots & f_{\text{PA},L}(z_L) \end{bmatrix}^T. \tag{C.2b}$$

The PA output $y_{\text{PA},l} = f_{\text{PA},l}(z_l)$ is given by passing a predistorted signal z_l through some nonlinear PA function $f_{\text{PA},l}(\cdot)$, which e.g. could be a Volterra-series, commonly known able to model nonlinear PAs. Further, to consider multiple directions simultaneously, OTA observations can be stacked into a vector

$$\boldsymbol{y} = \boldsymbol{H} f_{\text{PA}}(\boldsymbol{z}), \tag{C.3}$$

where

$$\boldsymbol{y} = \begin{bmatrix} y^{(1)} & \dots & y^{(Q)} \end{bmatrix}^T, \boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}^{(1)} \\ \vdots \\ \boldsymbol{h}^{(Q)} \end{bmatrix}. \tag{C.4}$$

Additionally, let each predistorted signal be generated by a DPD modelled as

$$z_l = f_{\text{DPD},l}(v_l) = \boldsymbol{F}_l \boldsymbol{\theta}_l, \tag{C.5}$$

where $\boldsymbol{F}_l$ are DPD bases and $\boldsymbol{\theta}_l$ modeling coefficients. Exemplified, these can be

$$\boldsymbol{F}_l = \begin{bmatrix} v_l & v_l|v_l| & \dots & v_l|v_l|^{P-1} \end{bmatrix}, \boldsymbol{\theta}_l = \begin{bmatrix} \theta_1 & \dots & \theta_P \end{bmatrix}^T, \tag{C.6}$$

for a P:th order static polynomial model.

4 Spatial and frequency weighted DPD

In identifying the optimal parameters of a set of DPDs, some criterion of optimality must be stated. In this section, the first subsection details the conventional criteria of minimizing the error at each PA output in relation to a desired output. In the second subsection, the proposed criterion of this paper is detailed, comparing OTA observation errors weighed differently across different regions in space and frequency.

4.1 Conventional SISO DPD

Conventionally, the transmitter of Fig. 1 is linearized through finding the DPD coefficients $\boldsymbol{\theta}_l$ for each branch $l = 1, ..., L$ that fulfill

$$\arg\min_{\boldsymbol{\theta}_l} ||y_{\mathrm{PA,d},l} - f_{\mathrm{PA},l}(\boldsymbol{F}_l\boldsymbol{\theta}_l)||^2 \,. \tag{C.7}$$

That is, the DPD coefficients are chosen to minimize the error from a desired $y_{\mathrm{PA,d},l}$ at each PA output. For this work, this approach will be denoted as per-PA DPD (PP-DPD). Typically, the desired PA output is a linear scaling of the precoded input v_l, as $y_{\mathrm{PA,d},l} = gv_l$. Two statements are worth to note in relation to this criterion. Firstly, for each amplifier, there is only one optimal solution to (C.7). Secondly, the criterion (C.7) does not account for the antennae, or channel effects.

4.2 Criteria for DPD optimization

If observing the signal OTA, instead of at the PA output, with the intent of finding a set of optimal DPD coefficients jointly across the branches, then a natural formulation would be

$$\arg\min_{\boldsymbol{\theta}} ||\boldsymbol{y}_{\mathrm{d}} - \boldsymbol{H} f_{\mathrm{PA}}(\boldsymbol{F}\boldsymbol{\theta})|| \,, \tag{C.8}$$

where $\boldsymbol{y}_{\mathrm{d}}$ is the set of desired signals in directions $q = 1, ..., Q$, of the same form as (C.3), and

$$\boldsymbol{F} = \begin{bmatrix} \boldsymbol{F}_1 & \boldsymbol{0} & ... & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & ... & \boldsymbol{F}_L \end{bmatrix}, \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_L \end{bmatrix} . \tag{C.9}$$

Here, each $\boldsymbol{F}_l$ denotes the DPD basis for branch l. A key difference between (C.7) and (C.8) is that in (C.8) the linear combination of signals OTA is optimized as the output. In (C.7), in contrast, the DPDs are optimized separately. The formulation in (C.9), with the use of a block matrix, can thereby be understood as necessary, since the DPDs across branches should be optimized jointly. Note that in (C.8) the linear combination over-the-air contains spatial weighting, dependent on the channel gain of $\boldsymbol{H}$. From this, distortion can be tolerated in directions with low channel gain without significantly affecting the optimality of $\boldsymbol{\theta}$. This can be further exploited by applying different weights in different spatial directions, corresponding to where an operator can tolerate distortion. To be described next is how IB and OOB distortion can be considered separately to obtain joint spatial and frequency weighting.

To obtain a frequency weighted formulation, consider that K baseband samples are processed. Let $\hat{\boldsymbol{F}}$ be the frequency-domain version of $\boldsymbol{F}$ from (C.9), obtained by Fourier-transforming each block $\boldsymbol{F}_l$. Similarly, let $\hat{\boldsymbol{y}}_\mathrm{d}$ denote the frequency domain representation of $\boldsymbol{y}_\mathrm{d}$. Introducing spatial-frequency weights $\boldsymbol{w}_{k,q}$ for the frequency indices k and spatial indices $q = 1, ..., Q$, then (C.8) can be extended to

$$\arg\min_{\boldsymbol{\Theta}} \sum_{q=1}^{Q} \sum_{k=1}^{K} \boldsymbol{w}_{q,k} \left| \hat{\boldsymbol{y}}_\mathrm{d}^{(q)}[k] - \boldsymbol{h}^{(q)} f_\mathrm{PA}(\hat{\boldsymbol{F}}[k]\boldsymbol{\Theta}^{[k]}) \right|. \tag{C.10}$$

Here, $\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}^{[1]} & ... & \boldsymbol{\Theta}^{[K]} \end{bmatrix}$ is a set of DPD coefficients to be applied at different frequencies k, where each $\boldsymbol{\Theta}^{[k]}$ is a vector as in (C.9). In contrast to (C.7), (C.10) can be underdetermined and yield nonunique solutions, unless $|w_{q,k}| > 0$ holds in at least L spatial directions. With $w_{q,k} = 1$ everywhere, and $\boldsymbol{\Theta}^{[k]} = \boldsymbol{\theta}\,, \forall k$, (C.10) is equivalent to (C.8).

Notably, (C.10) allows distortion to be shaped in both frequency and space by giving relative weight to where the distortion is most harmful to the system. Whilst (C.7) minimizes the distortion at each PA output, (C.10) allows for altering the DPDs to a solution better in some spatial and frequency regions, at the cost of degradation in other regions. This allows for flexibility in adhering to emission requirements, whilst obtaining improved performance in user directions.

Conventional DPD identification, i.e. solving (C.7), is a well-studied and understood problem, requiring no further details here. However, (C.10) being a novel criterion, the following section will detail approaches to identification.

5 Identification of the spatially and frequency weighted DPD

As a first note on solving (C.10), then doing so for K unique sets of DPD coefficients is most likely excessive. In practice, considering a few selected frequency regions $\boldsymbol{\kappa}_1, ..., \boldsymbol{\kappa}_S$ instead of the form

$$\boldsymbol{\Theta}^{[k]} = \boldsymbol{\theta}_i \quad \text{if}\, k \in \boldsymbol{\kappa}_i \,, i = 1, ..., S \,, \tag{C.11}$$

is likely a more feasible approach. The case $S = 2$ can correspond to IB and OOB separation, if the sets $\boldsymbol{\kappa}_1, \boldsymbol{\kappa}_2$ are chosen correspondingly.

As a second note, solving (C.10) can, as in the conventional case, be done by direct or indirect learning methods. Typically, either method is done in an iterative manner. So, regardless of approach, a suitable first step is forming the weighted OTA error, in the frequency domain, at iteration $[i]$, for each frequency component k and spatial direction q as

$$\hat{e}_{q,k}^{[i]} = \boldsymbol{w}_{q,k} \left| \hat{\boldsymbol{y}}_{\mathrm{d}}^{(q)}[k] - \boldsymbol{h}^{(q)} f_{\mathrm{PA}}(\hat{\boldsymbol{F}}[k]\boldsymbol{\Theta}^{[k]}) \right| . \tag{C.12}$$

Then, the errors are combined into a matrix $\hat{\boldsymbol{E}}^{[i]}$ of size $Q \times K$ with elements $\hat{e}_{q,k}^{[i]}$.

An indirect approach, such as iterative learning control (ILC) [11] can then solve for an optimal set of predistorted input signals as

$$\hat{\boldsymbol{z}}^{[i+1]} = \hat{\boldsymbol{z}}^{[i]} + \eta \boldsymbol{H}^{\dagger} \hat{\boldsymbol{E}}^{[i]} \,, \tag{C.13}$$

where η is a step size, and $\boldsymbol{H}^{\dagger} = (\boldsymbol{H}^{*}\boldsymbol{H})^{-1}\boldsymbol{H}^{*}$ denotes the pseudo-inverse of the channel matrix, after which DPD identification is performed. Alternatively, for direct learning methods, forming a DPD basis

$$\hat{\boldsymbol{F}} = \begin{bmatrix} \hat{\boldsymbol{F}}_1[\kappa_1] & \boldsymbol{0} & ... & \boldsymbol{0} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & ... & \hat{\boldsymbol{F}}_L[\kappa_S] \end{bmatrix} , \tag{C.14}$$

then allows for direct schemes, for example similar to the approach in [8], such that at iteration $[i+1]$

$$\begin{bmatrix} \boldsymbol{\theta}_{\kappa_1} \\ \vdots \\ \boldsymbol{\theta}_{\kappa_S} \end{bmatrix}^{[i+1]} = \begin{bmatrix} \boldsymbol{\theta}_{\kappa_1} \\ \vdots \\ \boldsymbol{\theta}_{\kappa_S} \end{bmatrix}^{[i]} + \eta \hat{\boldsymbol{F}}^{\dagger} \boldsymbol{H}^{\dagger} \hat{\boldsymbol{E}}^{[i]} \,. \tag{C.15}$$

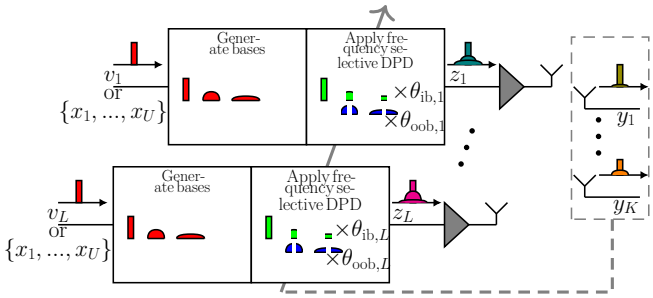


Figure 2: Spatially and frequency weighted DPD with two frequency regions, in-band (ib) and out-of-band (oob).

In (C.14), the bases can depend on the number of users, as will be detailed in the following subsections. The DPD using a basis as in (C.14), optimized for (C.10), constitutes a spatially and frequency weighted (SFW) DPD. This scheme is visualized in Fig. 2, and visualizes a case with $S = 2$ in relation to (C.11).

5.1 Single beam

Considering a single user beam, the distortion prior to linearization will be emitted in the same direction as the user. Then, the DPD objective would most often be to reduce and shape distortion away from that direction. This can be achieved through joint optimization of the DPD coefficients for a basis (C.14) formed from precoded inputs v_l, e.g. based on (C.6). This will result in that DPD coefficients are tapered along the branches, in relation to the conventional PP-DPD, which is later seen in Fig. 3.

5.2 Multiple beams

With beams to $U > 1$ users, there will be distortion going in off-beam directions, in addition to the user beam directions. An alternative is then to work with a beam-domain basis as in [12] for (C.14), i.e. based on input streams $x_1, ..., x_U$, exemplified with

$$\boldsymbol{F}_l = \begin{bmatrix} x_1 & ... & x_U & ... & x_1(|x_1|^2 + ... + |x_U|^2) & ... \end{bmatrix}, \tag{C.16}$$

Table 1: Distortion weights for different parts of the frequency spectrum and angular space. Larger weight magnitudes corresponds to that distortion there should be penalized more.

Distortion weights $w_{q,k}$ (Lin)	In-band	Out-of-band
Towards users $\{90 \pm 10\}^\circ$	5	4.25
Towards nonusers	0.01	0.5

(a) Weights corresponding to Fig. 3, a single-beam case focused on user-oriented distortion reduction.

Distortion weights $w_{q,k}$ (Lin)	In-band	Out-of-band
Towards users $\{76 \pm 5,\ 97 \pm 5\}^\circ$	1	5
Towards nonusers	0.01	5

(b) Weights corresponding to Fig. 4, a dual-beam case focused on out-of-band distortion reduction everywhere.

instead of $\boldsymbol{F}_l$ based on precoded inputs v_l. This gives greater flexibility in shaping the distortion based on each stream. The additional cost is that of multi-dimensional, rather than SISO, DPDs for such an implementation. The schematic of Fig. 1 is then altered such that $\boldsymbol{x}_1, ..., \boldsymbol{x}_U$ are input to the DPDs, as in Fig. 2, resulting in multi-input, single-output DPDs.

6 Simulation setup and results

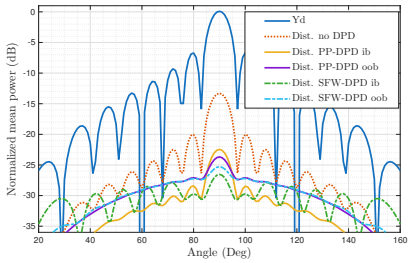
For simulation, the channel $\boldsymbol{H}$ is chosen as a line-of-sight (LOS) channel with antenna array pattern gain varying with angle ϕ as $\sin^3(\phi)$[1], and with $L = 16$ branches and $Q = 179$ observation directions. The PA function is implemented as a baseband Saleh model [13], equal on each PA branch, with added gaussian noise $\boldsymbol{n}_l$ at the output as

$$f_{\mathrm{PA},l}(\boldsymbol{z}_l) = \frac{a\boldsymbol{z}_l}{1 + b|\boldsymbol{z}_l|^2} + \boldsymbol{n}_l\,, \tag{C.17}$$

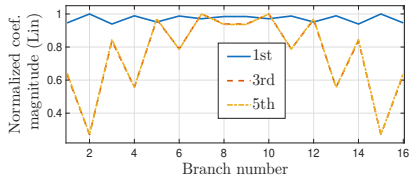
with $a = 1, b = 1$ and a standard deviation $\boldsymbol{\sigma}_{v_l} \approx 0.2$ of the precoded inputs $\boldsymbol{v}_l$. The desired signal is

$$\boldsymbol{Y}_{\mathrm{d}} = a_d \boldsymbol{H}\boldsymbol{v}\,, \tag{C.18}$$

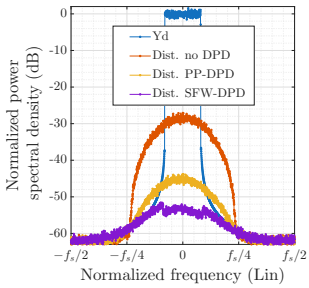
[1]The $\sin^3(\phi)$ function is a modeling choice, loosely based on experimental observations, to exemplify an array radiating worse towards the end-fire direction than towards broadside.



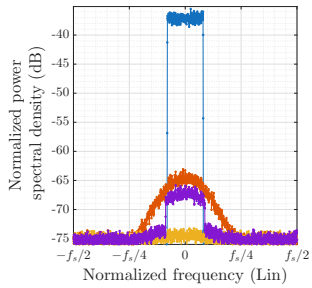
(a) Signal and distortion power, plotted separately for in-band (ib) and out-of-band (oob), in varying spatial directions.



(b) Normalized, relative magnitude of the first, third and fifth order in-band SFW-DPD coefficients. Coefficient tapering along the branches is seen, shaping distortion.



(c) Power spectral density in user direction, 90°.

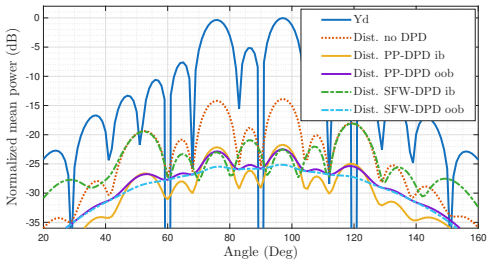


(d) Power spectral density in non-user direction, 62°.

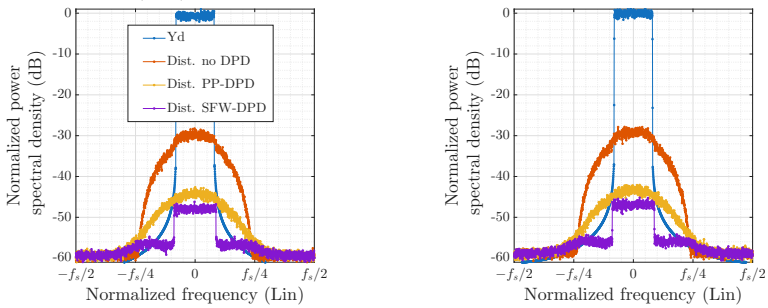
Figure 3: Normalized power in contrast to desired signal, for no DPD, conventional per-PA (PP) DPD or with spatially and frequency weighted (SFW) DPD, for a single beam scenario.

with $a_d = 0.88$ chosen as a desired linear gain of the PAs. Baseband data of $N = 65536$ samples is generated as bandlimited Gaussian noise with normalized bandwidth 0.16, allowing spectral regrowth.

In Fig. 3, a single-beam case with $U = 1$ user is shown. The proposed criterion (C.10) is optimized using direct DPD learning as in (C.15), with weights $w_{q,k}$ as listed in Table 1a. In Fig. 3a and Fig. 3c it can be seen that the error towards the user is decreased up to 8 dB by shaping IB distortion in other spatial directions, for SFW-DPD compared to conventional PP-DPD. Fig. 3b illustrates how this SFW-DPD solution involves tapering of the DPD coefficients to achieve the distortion shaping. The weights of Table 1a are chosen, somewhat arbitrarily, to reflect higher importance of especially in-



(a) Signal and distortion power, plotted separately for in-band (ib) and out-of-band (oob), in varying spatial directions.



(b) Power spectral density in direction 76°. **(c)** Power spectral density in direction 97°.

Figure 4: Normalized power in contrast to desired signal, for no DPD, conventional per-PA (PP) DPD or with spatially and frequency weighted (SFW) DPD, for a dual-beam scenario.

band user-oriented distortion. Larger weight magnitudes corresponds to that distortion there should be penalized more. The DPDs, modelled based on (C.6), include odd and even order static polynomial terms up to nonlinearity order $P = 7$.

In Fig. 4, a multi-beam case with $U = 2$ users is explored, with $w_{q,k}$ as in Table 1b. From Fig. 4a and Fig. 4b it is seen that OOB distortion is reduced everywhere in space for the proposed SFW-DPD compared to PP-DPD, at the expense of increased IB distortion in non-user directions. The total emitted OOB power is reduced by approximately 2.5 dB, whilst user-directed IB distortion is simultaneously lowered. The weights of Table 1b are, in contrast to Table 1a, chosen to reflect higher importance of out-of-band distortion emitted overall. The DPDs, based on (C.16), include odd and even order static polynomial terms up to nonlinearity order $P = 5$.

7 Discussion

Implementing DPDs per branch is costly in large MIMO transmitters, which has motivated other work in e.g. beam-domain DPD [12]. And whilst the proposed work is not mainly focused on complexity reduction, complexity concerns are worth further comment. Firstly, one way to regulate complexity in the proposed DPD is to only linearize a subset of the frequency indices. Secondly, DPDs linearizing different frequency regions can include different number of coefficients, as the strictness of linearization requirements can vary. Lastly, an alternative approach to complexity reduction is to combine the proposed SFW-DPD with other complexity reduction methods such as [14], where the number of DPDs L is reduced through a so-called virtual array DPD.

These combined strategies suggests that the SFW-DPD can be adapted to meet various implementation constraints.

8 Conclusion

Distortion shaping in both frequency and space is shown capable of improving linearization performance in specified frequency and spatial regions at the cost of shaping distortion elsewhere. The proposed error metric allows flexibility in how residual distortion is to be shaped, and shown implementable by the proposed spatial and frequency weighted (SFW) DPD approach. Specifically, simulation results for single-beam and dual-beam scenarios illustrate the ability of the SFW-DPD to reduce user-oriented distortion, selectively suppress out-of-band emissions, and exploit spatial and spectral degrees of freedom to meet diverse emission requirements. Overall, the approach offers a flexible framework for distortion management in digital MIMO transmitters.

References

[1] N. Kolomvakis, M. Bavand, I. Bahceci, and U. Gustavsson, "A Distortion Nullforming Precoder in Massive MIMO Systems With Nonlinear Hardware," *IEEE Wireless Communications Letters*, vol. 11, no. 9, pp. 1775–1779, Sep. 2022, ISSN: 21622345.

[2] S. R. Aghdam, S. Jacobsson, U. Gustavsson, G. Durisi, C. Studer, and T. Eriksson, "Distortion-Aware Linear Precoding for Massive MIMO Downlink Systems with Nonlinear Power Amplifiers," Nov. 2021.

[3] N. Tervo, J. Aikio, T. Tuovinen, T. Rahkonen, and A. Parssinen, "Digital predistortion of amplitude varying phased array utilising over-the-air combining," in *IEEE MTT-S International Microwave Symposium*, IEEE, 2017, pp. 1165–1168.

[4] N. Tervo, B. Khan, J. P. Aikio, *et al.*, "Combined Sidelobe Reduction and Omnidirectional Linearization of Phased Array by Using Tapered Power Amplifier Biasing and Digital Predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 9, pp. 4284–4299, Sep. 2021, ISSN: 15579670.

[5] J. Yan, H. Wang, and J. Shen, "Novel Post-Weighting Digital Predistortion Structures for Hybrid Beamforming Systems," *IEEE Communications Letters*, vol. 25, no. 12, pp. 3980–3984, Dec. 2021, ISSN: 15582558.

[6] Z. Zhang, Y. Li, S. Yang, N. Wei, and C. Yuen, "Linearization Angle Widened Predistortion for Hybrid Beamforming Array Utilizing Iterative Post-Weighting," in *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, May 2025, pp. 1–5, ISBN: 979-8-3503-5683-0.

[7] X. Liu, Q. Zhang, W. Chen, *et al.*, "Beam-Oriented Digital Predistortion for 5G Massive MIMO Hybrid Beamforming Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018, ISSN: 00189480.

[8] A. Brihuega, L. Anttila, and M. Valkama, "Frequency-Domain Digital Predistortion for OFDM," *IEEE Microwave and Wireless Components Letters*, vol. 31, no. 6, pp. 816–818, Jun. 2021, ISSN: 15581764.

[9] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited volterra series-based digital predistortion for wideband RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 4198–4208, 2012, ISSN: 00189480.

[10] H. Yin and A. Zhu, "Pareto-Optimal Multimetric Model Extraction for Digital Predistortion of RF Power Amplifiers for Error Spectrum Redistribution," *IEEE Transactions on Microwave Theory and Techniques*, vol. 73, no. 4, pp. 2230–2241, 2025, ISSN: 15579670.

[11] J. Chani-Cahuana, P. N. Landin, C. Fager, and T. Eriksson, "Iterative Learning Control for RF Power Amplifier Linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 9, pp. 2778–2789, Sep. 2016, ISSN: 0018-9480.

[12] A. Brihuega, L. Anttila, and M. Valkama, "Beam-Level Frequency-Domain Digital Predistortion for OFDM Massive MIMO Transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 4, pp. 1412–1427, 2023.

[13] A. Saleh, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *IEEE Transactions on Communications*, vol. 29, no. 11, pp. 1715–1720, 1981.

[14] B. Langborn, C. Fager, R. Hou, and T. Eriksson, "Concurrent multi-beam digital pre-distortion using FFT beamforming and virtual arrays," *Preprints, February 2026, doi: 10.20944/preprints202602.0713.v1*,