Signal Theory and Communications Department

Universitat Politècnica de Catalunya

Ph.D. Thesis

Machine Learning Techniques for Adaptive Polynomial and Neural Network Digital Predistorters

Author
David Lopez Bueno

Supervisors
Dr. Pere L. Gilabert Pinal
Dr. Gabriel Montoro López

Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Phylosophy in Signal Theory and Communications
at the Universitat Politècnica de Catalunya

Barcelona
May, 2023

To my family,

In the world of digital communication,
Where wireless signals fill the air,
There lies a field of innovation,
Where machine learning techniques repair.

Polynomial and neural network predistorters,
Adaptive and complex they may be,
But with machine learning as their supporters,
Their computational load is set free.

Reduced complexity and DPD training times,
Are the gifts that machine learning brings,
A boon for the engineers and their rhymes,
And the efficiency of wireless things.

Power amplifiers, the heart of wireless might,
Can suffer from distortion and nonlinearity,
But with digital predistortion as their guiding light,
Their efficiency and performance can reach clarity.

And for those who seek knowledge and insight,
In this field of beauty and innovation,
Thesis writing can bring great delight,
As they uncover new ways of predistortion.

So let us honor the beauty and the gains,
Of machine learning in the digital realm,
Where innovation and efficiency reigns,
And the future is at the helm.

Poem on the Beauty and Benefits of Machine Learning Techniques in Digital Predistorters, Generated by ChatGPT, a large language model trained by OpenAI, based on the GPT-3.5 architecture.

The author declares this poem is strictly the only use of chatGPT made over the whole thesis manuscript. May 25th, 2023.

Abstract

The power amplifier (PA) is a core element in the radio transmitters to support the required mobile and fixed broadband communication link ranges. However, the PA is a power-hungry and nonlinear by nature device. Under spectrally efficient wideband modulated waveforms with high peak-to-average power ratio, the PA energy efficiency is significantly decreased since back-off operation is needed to meet the transmission quality requirements. Moreover, when employing highly efficient amplification architectures, like Doherty, load modulated balanced amplifier (LMBA) or envelope tracking (ET) PAs, the added distortion is left as an issue to be addressed at system level by a linearizer. In this context, the closed-loop adaptive digital predistorter (DPD) is a key component of the digital front-end (DFE) to counteract the PA nonlinear response under varying conditions and to cope with the inherent trade-off between linearity and efficiency.

According to the fifth generation (5G) and beyond communication technologies and the proposed radio transmitter and PA architectures, the DPD may have to deal with strong nonlinearities and memory effects, in-phase and quadrature (IQ) modulator imbalances and dc offsets, additional PA supply or load modulation distortion, and multi-antenna PA input and output cross talk and beam-dependent effects. Such impairments degrade the radio access network (RAN) energy efficiency, capacity, and the number of potential RAN users due to the increased in-band and out-of-band distortion. The adaptive DPD can overcome such undesired effects but faces relevant obstacles. At every new generation of mobile communication systems, the signal bandwidth is increased and the DPD needs to be operated at higher speed. The DPD challenges are twofold. On the one hand, combining massive bandwidth operation together with handling complex multi-dimensional effects may increase exponentially the complexity of the DPD and make it both commercially unaffordable and energy inefficient due to the increased cost and power consumption at the DFE and data conversion stages. On the other hand, the adaptive DPDs need significantly larger training periods to compensate for all the complex undesired effects.

In line with the above-mentioned challenges, the research presented in this dissertation aims at guaranteeing best DPD linearity versus efficiency trade-off in complex nonlinear scenarios, by leveraging on efficiently deployed machine learning (ML) and artificial intelligence (AI) techniques to reduce the computational complexity of DPD modeling and identification at the DFE, guaranteeing well-conditioned and robust DPD estimation, and drastically reducing the DPD training times while meeting performance requirements. To accomplish that, several newly applied and customized ML feature selection and feature extraction dimensionality reduction techniques are combined with new training data length reduction schemes, to ensure both reduced DPD behavioral modeling matrices and input datasets in single-antenna and multi-antenna adaptive polynomial and neural network DPD architectures, respectively. To validate the benefits of these contributions in accordance with the previous goals, all these techniques have been deployed and thoroughly benchmarked under adverse conditions in realistic laboratory test benches.

Resum

L'amplificador de potència (PA) és un element central als radiotransmissors necessari per abastar les distàncies d'enllaç a les comunicacions sense fils de banda ampla mòbil i fixa. No obstant, el PA és un dispositiu d'elevat consum energètic i de comportament no lineal per naturalesa. Quan s'utilitzen formes d'ona de banda ampla i gran eficiència espectral que presenten elevada relació de potència de pic a potència mitjana, l'eficiència energètica del PA decreix significativament atès que és necessari operar el dispositiu amb un marge de guarda per complir amb els requisits de qualitat de transmissió. Addicionalment, quan s'utilitzen arquitectures d'amplificació altament eficients basades en PAs de tipus Doherty, modulació de càrrega d'amplificador balancejat (LMBA) o modulació d'envoltant (ET), la distorsió afegida es deixa com un problema a resoldre a nivell de sistema per un linealitzador. En aquest context, el predistorsionador digital (DPD) adaptatiu de llaç tancat és un element clau del capçal digital de processament ràdio (DFE) per contrarestar la resposta no lineal del PA en condicions variables i superar el compromís inherent entre linealitat i eficiència.

D'acord amb les tecnologies de comunicació de cinquena generació (5G) i les arquitectures de transmissor i PA proposades, el DPD pot haver de tractar amb fortes no linealitats i efectes de memòria, desajustaments de modulador en fase i quadratura (IQ) i desviacions de nivell de contínua, distorsió addicional per modulació d'alimentació o càrrega del PA, efectes d'acoblament creuat a l'entrada i sortida dels PAs en sistemes multiantena, així com efectes que poden dependre de l'apuntament dels feixos de l'antena. Aquestes imperfeccions degraden l'eficiència de la xarxa d'accés ràdio (RAN), la seva capacitat i el nombre d'usuaris a causa de l'augment de distorsió generada dins de banda i fora de banda. El DPD adaptatiu pot combatre aquests efectes no desitjats però s'enfronta a obstacles rellevants. A cada nova generació de sistemes de comunicacions mòbils l'amplada de banda de senyal s'incrementa i el DPD necessita operar a majors velocitats. Els reptes imposats al DPD són doblement exigents. D'una banda, combinar amplades de banda de senyal massives amb haver de modelar efectes complexes multidimensionals pot augmentar exponencialment la complexitat del DPD i fer-lo comercialment inviable i energèticament ineficient donat l'increment en cost i consum al DFE i a les etapes de conversió de senyal. D'altra banda, els DPD adaptatius necessiten períodes d'entrenament significativament més grans per compensar tots els efectes complexes no desitjats.

D'acord amb els reptes esmentats, la recerca presentada en aquesta dissertació té com a objectiu garantir el millor compromís entre linealització del DPD i eficiència assolida en escenaris amb no linealitats complexes, mitjançant tècniques eficientment desplegades d'aprenentatge màquina (ML) i d'intel-ligència artificial (AI) necessàries per reduir la complexitat computacional del modelat i la identificació DPD al DFE, garantir una estimació DPD ben condicionada i robusta, i reduir dràsticament els temps d'entrenament del DPD, sense deixar d'acomplir els requisits de qualitat de transmissió. Per aconseguir-ho, en aquesta tesi s'apliquen de forma innovadora tècniques adaptades de reducció de dimensionalitat ML basades en selecció i extracció de característiques (features), i es combinen amb nous esquemes de reducció de llargada de les dades d'entrenament DPD, per garantir tant la reducció de les matrius de modelat comporta-

mental DPD polinomial com els datasets d'entrada a predistorsionadors adaptatius basats en xarxes neuronals, tant per arquitectures monoantena com multiantena. Per validar els beneficis d'aquestes contribucions d'acord amb els objectius anteriors, totes aquestes tècniques s'han desplegat i avaluat minuciosament sota condicions adverses en bancs de proves de laboratori propers a condicions reals.

Acknowledgements

This thesis was carried out part-time at the Signal Theory and Communications Department (TSC) at Universitat Politècnica de Catalunya (UPC), while working full-time at the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC). Most of the works presented in this thesis were conducted in the frame of research projects where the two institutions were collaborating. These projects were financially supported in part by Spanish, Catalan and European grants covering the cost of the equipment and parts needed for the experimental test setups, conference registration and journal publication fees, and the travel expenses. The author would like to acknowledge the Spanish Government (including the Ministerio de Ciencia, Innovación y Universidades and the Ministerio de Asuntos Económicos y Transformación Digital), and the Fondo Europeo de Desarrollo Regional under projects TEC2014-58341-C4-3-R, TEC2014-58341-C4-4-R, TEC2017-83343-C4-2-R, RTI2018-099841-B-I00, PID2020-113832RB-C21 and PID2020-113832RB-C22 (MCIN/AEI/10.13039/501100011033), TSI-063000-2021-121 (MINECO UNICO 5G I+D programme funded by the EU NextGenerationEU programme through the Spanish Recovery, Transformation and Resilience Plan). Also, the Generalitat de Catalunya (Universities and Research Department) under grants 2014 SGR 1551, 2017 SGR 813, 2017 SGR 891, 2017 SGR 1479, and 2021 SGR 00772. Finally, the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement 101096034.

First and foremost, I would like to express my gratitude to my thesis supervisors Prof. Pere L. Gilabert and Prof. Gabriel Montoro for their invaluable guidance and support throughout all the thesis years. I feel honored for having had the opportunity to learn, do research and work in projects with so knowledgeable, creative, passionate, hands-on enthusiasts, tireless workers and friendly researchers. Thanks also for pushing me out of my comfort zone so many times and your great willingness to help, also for accepting the turment to read my tough manuscripts (you know I am a failed novel writer, but chatGPT will help me for sure to be more concise, or maybe not). I am grateful for having shared so many good moments with you over these years, also to have met my PhD pals Quynh Anh Pham, Teng Wang and Wantao Li. I am very happy to have collaborated in research works that have also contributed to enrich this thesis, and having had fun both in the lab and out of it. I wish you all the best, the future is yours!

I also wish to thank the external evaluators Prof. Roberto Quaglia and Prof. Maria José Madero Ayora, for their availability and outstanding responsiveness, for all the comments provided that have contributed to polish the thesis manuscript and improve its quality. Also, for accepting being part of the Thesis Defense Committee, together with Prof. Jordi Mateu Mateu who I am also grateful to count with. In addition, I would like to thank our colleagues Prof. José A. García and Dr. Nieves Ruiz from the Universidad de Cantabria for the donation of the two wonderful class-J PA that have been used 24x7 in numerous demanding experiments over

long periods, always demonstrating high reliability and a consistent robust performance and facilitating the benchmarks presented in this Thesis. Working with you in projects and meeting at conferences is always a pleasure.

At CTTC side, I seize the occasion to thank the Director Prof. Ana Pérez, the former Director Prof. Miguel Ángel Lagunas, and the General Administrator Ms. Mercè Carrasco, for giving me the opportunity to be part of one of the forefront telecommunication and geomatics research centers in Europe, continuously promoting the pursuit of scientific excellence in a collaborative space, encouraging engineers to get a PhD, and working towards a high-quality workplace environment. I also extend my thanks to Dr. Miquel Payaró former Head of the Communication Technologies Division, now Director of Open Innovation and Science.

I want also to express my sincere gratitude to Dr. Nikolaos Bartzoudis, Head of the Adaptive Processing Technologies (ADAPT) Research Unit, for placing his trust in me and in a new research line linked to the thesis topics that has enabled numerous collaborations in projects and publications. After all these years working and interacting with you Niko, I am certain that this thesis will also have a part of your imprint. Together with Nikos, I want also to thank other colleagues like Oriol F. and Pepe R. who have been contributing to the real-time implementation of digital linearization systems in projects, and have also strengthened the research line.

I cannot miss the opportunity to thank some of the CTTC colleagues (Coffee Club & Lunchtime Crime Syndicate Members) for enhancing my mood at work, my general knowledge, and the insights on friki-nerd-geek culture. Especially, those with whom I have shared many things inside and outside the CTTC like David P., Marco M., Xavi A., O. Font and Paolo D. (ex-Burning Transistors), Luis S., Javier V. and our last recruit Antonio (Hulk) R. I extend my thanks to other researcher and non-researcher colleagues I interact with day by day.

Now it is the turn of several long-life friends like Fredi, Josep, Rubén, Joan and Sílvia, Xavi. This thesis closes some period in my life where I have not found the chance to meet with you as frequently as I would have liked. Let's fix it!

Cristina, gràcies per posar llum a la meva vida, per suportar els meus horaris o canvis d'humor en determinats moments d'aquesta tesi, per estar sempre al meu costat i ensenyar-me què és l'amor incondicional. No puc expressar amb paraules el que sento per tu i quant feliç em fas quan estic al teu costat. Aquesta tesi ha estat possible en part gràcies a tú també!

Finalmente, esta tesis la dedico a mis queridos padres Manuel y Mercedes, por despertar en mi la curiosidad desde temprana edad, por la educación que me habéis dado, por ser un referente de trabajo y esfuerzo en mi vida difícilmente igualable, por vuestra confianza, no puedo agradecer suficientemente todo lo que habéis hecho por mí. También a mi querida hermana Ruth.

David Lopez Bueno

Barcelona, May 2023.

Contents

Acronyms and Abbreviations

3GPP	Third Generation Partnership Project
4G	Fourth Generation (of mobile telecommunications technology)
5G	Fifth Generation (of mobile telecommunications technology)
6G	Sixth Generation
ACE	Active Constellation Extension
ACEPR	Adjacent Channer Error Power Ratio
ACPR	Adjacent Channel Power Ratio
ADC, A/D	Analog-to-Digital Converter
AEI	Agencia Estatal de Investigación
AI	Artificial Intelligence
AM-AM, AM/AM	Amplitude Modulation to Amplitude Modulation
AM-PM, AM/PM	Amplitude Modulation to Phase Modulation
ANN	Artificial Neural Network
APCA	Adaptive Principal Components Analysis
ASIC	Application-Specific Integrated Circuit
ATT	Attenuator/Attenuation
B5G	Beyond Fifth Generation
BB	Base Band
BB-PPDR	Broadband Public Protection and Disaster Relief (radio communications)
BBU	Baseband Unit
BD-APCA	Block-Deflated Adaptive Principal Components Analysis
BER	Bit Error Rate
BFG	Broyden-Fletecher-Goldfarb-Shanno
BM	Behavioral Modeling
BoM	Bill of Materials
BPA	Balanced Power Amplifier
BPF, LPF	Band-pass Filter, Low-pass Filter

BPLA	Back-propagation Learning Algorithm
bps	bits per second
BRAM	Block Random-Access Memory
BS	Base Station
BTS	Base Transceiver Station
BW	Bandwidth
CAPEX	Capital Expenditures
CBS	Consecutive Batch Selection
CCA	Contiguous Carrier Aggregation or Canonical Correlation Analysis (dep. on context)
CDF	Cumulative Distribution Function
CFG	Configuration
CFR	Crest Factor Reduction
CGB	Conjugate Gradient Backpropagation with Powell-Beale
CGF	Conjugate Gradient Backpropagation with Fletcher-Powell
CGHA	Complex domain Generalized Hebbian Algorithm
CGNAT	Carrier Grade Network Address Translation
CGP	Conjugate Gradient Backpropagation with Polak Ribiere
CGS, MGS	Classical/Modified Graam-Schmidt
CH	Channel
CLB	Configurable Logic Block
CLK	Clock
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Networks
CO2	Carbon dioxide
CORDIC	Coordinate Rotation Digital Computer
CP	Cyclic Prefix
CSP	Control Signal Power (amplifier)
CTTC	Centre Tecnològic de Telecomunicacions de Catalunya
CV	Complex-valued
CW	Continuous Wave
CX	Crosstalk
D2D	Device-to-device
DAC, D/A	Digital-to-Analog Converter
dB	decibel
dc	direct current (DC at start of sentence or in article/section title)

DDR	Dynamic Deviation Reduction
DF	Data Function
DFE	Digital Front End
DL	Downlink or Direct Learning (dep. on context)
DMO	Direct Mode Operation
DOMP	Doubly Orthogonal Matching Pursuit
DOTM	Dynamic Orthonormal Transformation Matrix (algorithm)
DPD	Digital Predistortion
DPLS	Dynamic Partial Least Squares (algorithm)
DRV	Driver
DSO	Digital Storage Oscilloscope
DSP	Digital Signal Processor
DTT	Digital Terrestrial Transmission
DUC/DDC	Digital Up/Down Converter
DUT	Device Under Test
DVB-T	Digital Video Broadcasting - Terrestrial
DVGA	Digital Variable Gain Amplifier
EC	European Commission
EER	Envelope Elimination and Restoration
EETAC	Escola d'Enginyeria de Telecomunicació i Aeroespacial de Castelldefels
e/gNodeB	Evolved/Next Generation Node B
ET-DPD	Envelope Tracking Digital Predistortion
EVM	Error Vector Magnitude
FC-FBMC	Fast Convolution Filter-Bank Multi-Carrier
FDD	Frequency Division Duplexing
FF	Flip Flop
FFT / IFFT	Fast Fourier Transform / Inverse Fast Fourier Transform
FIR	Finite Impulse Response
FFNN	Feedforward Neural Network
FPGA	Field-Programmable Gate Array
FTDNN	Focused Time-Delay Neural Network
FTP	File Transfer Protocol
GaAs	Gallium Arsenide
GaN	Gallium Nitride
GDA	Gradient Descent Algorithm
GDM	Gradient Descent with Momentum

GDX	Gradient Descent with adaptive learning and momentum
GFDM	Generalized Frequency Division Multiplex
GMP	Generalized Memory Polynomial
GMPNLC	Generalized Memory Polynomial for Nonlinear Crosstalk
GOH	Genetically Optimized Histogram
H-CRAN	Heterogeneous Cloud Radio Access Network
H2M	Human-to-Machine
HD	High-Definition
HEMT	High-Electron-Mobility Transistor
HetNet	Heterogeneous Network
HLS	High-Level Synthesis
HPA	High Power Amplifier
HPD	Hybrid predistorter/predistortion
HW	Hardware
Hz	Hertz
IAB	Integrated Access and Backhauling
IBO	Input Back-off
IBPR	In-band channel Power Rejection
ICI	Intercarrier Interference
ICT	Information and Communication Technologies
IEEE	Institute of Electrical and Electronics Engineers
IF	Intermediate Frequency
IL	Indirect Learning
IMD3	Third-Order Intermodulation Distortion
IMD5	Fifth-Order Intermodulation Distortion
IMS	International Microwave Symposium
IMT	International Mobile Telecommunication
IoT	Internet of Things
IQ, I/Q	In-phase and quadrature-phase
ITU	International Telecommunication Union
JEDEC	Joint Electron Device Engineering Council
KPI	Key Performance Indicator
LASSO	Least Absolute Shrinkage and Selection Operator
LINC	Linear amplification using Nonlinear Components
LM	Levenberg-Marquardt
LMBA	Load Modulated Balanced Amplifier

LMS	Least Mean Squares
LNA	Low Noise Amplifier
LO	Local Oscillator
LoS	Line-of-Sight
LRBR	Less Relevant Basis Removal
LS	Least Squares
LSTM	Long Short-term Memory
LTE	Long Term Evolution
LTI	Linear Time Invariant (system)
LUT	Look-up Table
M2M	Machine-to-Machine
MCIN / MICINN	Ministerio de Ciencia e Innovación
MCU	Minimum Computation Unit
MEC	Mobile Edge Computing
MINECO	Ministerio de Asuntos Económicos y Transformación Digital
MISO	Multiple-Input Single-Output
ML	Machine Learning
MLP	Multilayer Perceptron
mMIMO	massive Multiple-Input Multiple-Output
MOD / DEMOD	Modulation/Demodulation
MP	Memory Polynomial
MPA	Medium Power Amplifier
MSE	Mean Square Error
MTBF	Mean Time Between Failures
MTT-S	Microwave Theory and Techniques Society
MeS	Mesh Selecting
mmW, mm-wave	millimeter wave
NARMA	Non-Linear Auto-Regressive Moving Average
NCCA	Non-Contiguous Carrier Aggregation
NL	Nonlinear
NLoS	Non-Line-of-Sight
NMSE	Normalized Mean Square Error
NR	New Radio
NW	Network
N/A	Not Applicable
OFDM	Orthogonal Frequency Division Multiplexing

OLS	Orthogonal Least Squares
OMP	Orthogonal Matching Pursuit
OoB	Out-of-band
OPEX	Operating Expenses
OQAM	Offset Quadrature Amplitude Modulation
OSS	One Step Secant
P1dB	Output power when the amplifier is at the 1 dB compression point
PA	Power Amplifier
PAE	Power Added Efficiency
PAPR	Peak-to-Average Power Ratio
PC-CFR	Peak Cancellation Crest Factor Reduction
PCA	Principal Components Analysis
PHY	Physical (layer)
PL	Programmable Logic
PLS	Partial Least Squares
PS	Processing System
PS-LTE	Public Safety Long Term Evolution
Psat	Output power at when the amplifier is saturated
PSN	Public Safety Networks
PtMP	Point-to-MultiPoint
PtP	Point-to-Point
PTS	Partial Transmit Sequence
PWM	Pulse Width Modulation
PWR	Power
PXI	PCI eXtensions for Instrumentation
QAM	Quadrature Amplitude Modulation
QoS	Quality-of-Service
RAN	Radio Access Network
RBFNN	Radial Basis Function Neural Network
RF	Radio Frequency
RFIC	Radio Frequency Integrated Circuit
RLC	Radio Link Control
RLS	Recursive Least Squares
RNN	Recurrent Neural Network
RP	Resilient back-propagation
RR	Ridge Regression

RTL	Register Transfer Level
RV	Real-valued
RVTDNN	Real-Valued Time Delay Neural Network
RWW	Radio Wireless Week
RX / Rx	Receiver / Receive
Sa/s	Samples per second
SAW	Surface Accoustic Wave
SBL	Sparse Bayesian Learning
SBS	Sparse Batch Selection
SCG	Scaled Conjugated Gradient
SFB	Synthesis Filter Bank
SGR	Suport als Grups de Recerca
SHAPER	Solutions for wideband Highly lineAr and efficient PowER amplification
SiGe	Silicon Germanium
SINR	Signal-to-Interference and Noise Ratio
SISO	Single-Input Single-Output
SLM	Selected Mapping
SLP	Single Layer Perceptron
SMT	Staggered Multitone
SNR	Signal-to-Noise Ration
SoC	System-on-Chip
SOI	Silicon on Insulator
SotA	State-of-the-Art
SSM	Sample Selection Method
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
SW	Software
SW-OMP	Spectral Weighting Orthogonal Matching Pursuit
TCO	Total Cost of Ownership
TDD	Time Division Duplexing
TDL	Time Delay Line
TETRAPOL	Terrestrial Trunked Radio POLice
TI / TR	Tone Injection / Tone Reservation
TSC	Dept. de Teoria del Senyal i Comunicacions
TX / Tx	Transmitter / Transmit

uC	micro Controller
UE	User Equipment
UES	Uncorrelated Equation Selection
UFMC	Universal Filtered Multicarrier
UL	Uplink
UPC	Universitat Politècnica de Catalunya
Vds	Drain-to-source Voltage
VGA	Variable Gain Amplifier
Vgs	Gate-to-source Voltage
VST	Vector Signal Transceiver
W	Watt
WFM	Waveform
WRC	World Radiocommunication Conference

Chapter 1

Introduction

1.1 Motivation and Scope

1.1.1 Context

The exponential growth of the mobile broadband data requirements since the irruption of new mobile devices and applications has posed several technical and economic challenges to network equipment providers and telecom operators. While the network infrastructure technology has evolved to accommodate the higher capacity demanded by the users at an acceptable quality of service, the revenue growth has no longer followed the data traffic expansion [Möl10]. Higher investment in network deployment, increased operational costs, the decoupling between end user application value and the bit volume and the business monetization sharing with third-party service and application providers, are factors impacting into this reduced profitability.

Figure 1.1: Visual representation of the mobile data growth due to smartphone adoption and the deployment and launch of fourth-generation (4G) mobile networks and services [Tay13].

Reference studies provided by the industry are showing that every five years the mobile broadband data traffic scales by a factor of ten. The information and communication tech-

nologies (ICT) power consumption also scales by a factor of two every five years and its CO2 emission is approximately equivalent to that of airplanes and 1/4 that of automobiles worldwide [Kha14, Rit20], while the cost of producing the energy is also increasing on average. Spectrum efficiency is clearly no longer the only design criteria since energy efficiency represents a significant opportunity to reduce the total cost of ownership (TCO) of network operators and contribute to reduce the steadily increasing ICT carbon footprint [Fre21].

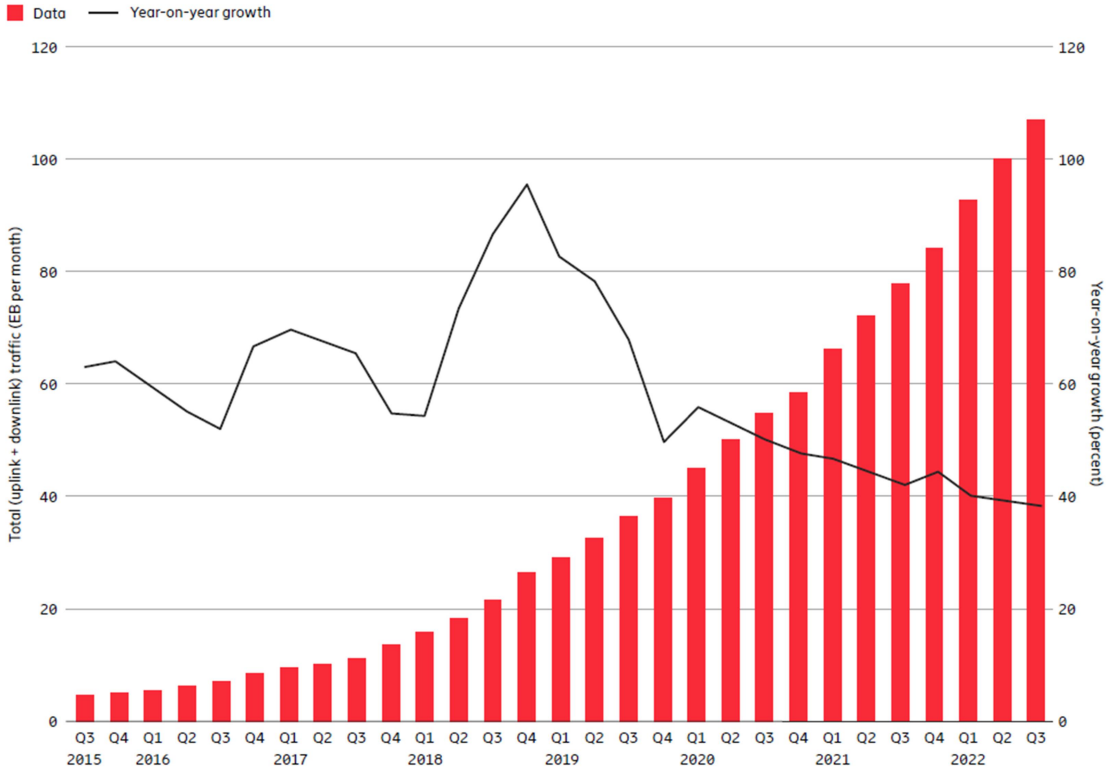


Figure 1.2: Global mobile network data traffic and year-on-year growth (EB per month, source: [Eri22]).

To try to minimize the power consumption in mobile networks and its carbon footprint, while having to increase their capacity, some general approaches are followed: The first consists in using energy efficient hardware implementations either in the analog or the digital signal processing domains targeting the best achievable trade-off between performance and power consumption, plus additional control software to adapt power consumption to the traffic situation. The second is based in intelligent network deployment strategies which typically place the radio access network (RAN) closer to the end user through smaller cells (with either smaller base stations or distributed base stations with remote radio head capability) to reduce the high transmit power energy implications and increase the capacity at the expense of demanding effective spectrum reuse and interference mitigation techniques and imposing further stringent backhaul equipment requirements. This approach is shown in the heterogeneous cloud CRAN (H-CRAN) architecture depicted in Fig. 1.3, which allows also distributing the classical base station or e/gNodeB

(e/gNB) building blocks by flexibly grouping PHY layer, MAC and RLC functions into different logical nodes or hardware (i.e., radio unit, distributed unit, centralized unit) according to the network requirements [Mio18]. As seen in [Bar17b], by flexibly distributing the processing load in reconfigurable splits it is possible to obtain energy savings in the radio unit. Finally, a third approach gaining attraction is also the use of renewable energy sources for RAN infrastructure both on-grid (powered from solar, wind and water sources) and off-grid deployments powered through on-site photovoltaic panels. This approach is combined with smart litium batteries and liquid cooling to reduce the need for air conditioning.

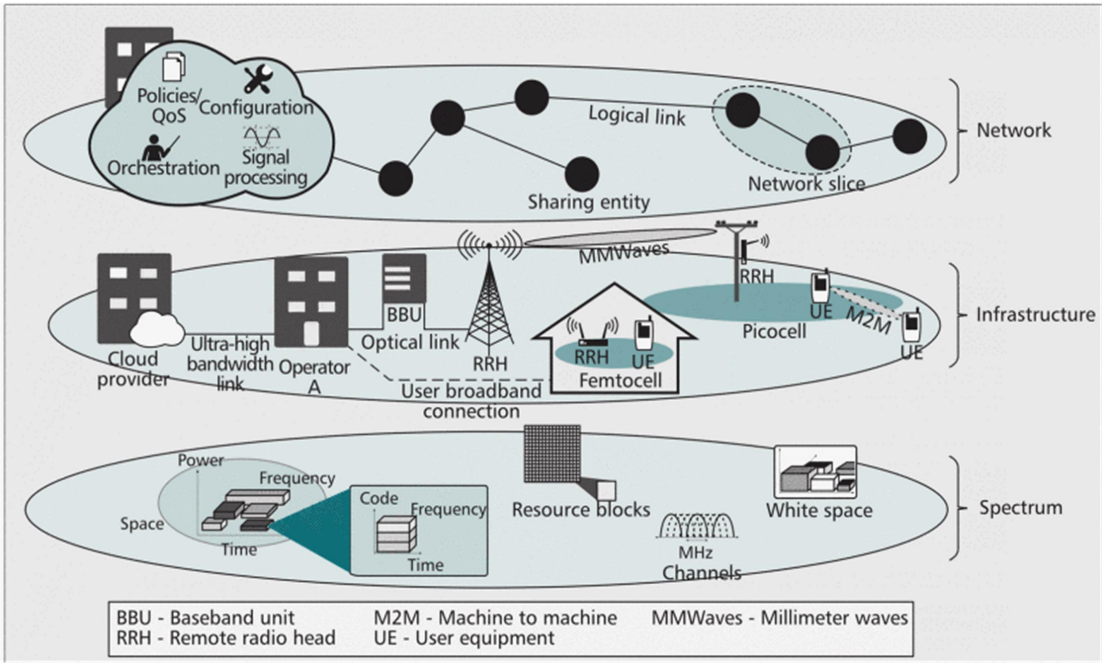


Figure 1.3: Resource sharing levels in H-CRAN architecture (source: [Mar15]).

In order to put the thesis research work in its technical context, the RAN and the radio equipment building blocks need to be introduced. The RAN may be defined as the main infrastructure used by the users in mobile broadband data networks covering the radio communication between the user equipment (UE, i.e., smartphones, mobile broadband modems, customer premises equipment, etc.) and the cellular base stations (BS), and the backhaul network connecting the base stations to the core network deploying secure connectivity and communication services to end users. Such basic components are shown in Fig. 1.4. The block diagram of a traditional 4G frequency division duplexing (FDD) base station or base transceiver station (BTS) is shown in Fig. 1.5 for illustrative purposes to introduce the most relevant radio equipment building blocks related to the scope of this thesis (i.e., appearing within the figure's yellow outline rectangles).

The 4G signals are modulated and demodulated in the "Digital Baseband" block or baseband unit (BBU) that in 4G is typically hosted in an equipment shelter beside the bottom of the base station tower. These digital signals, compliant to the air interface standard requirements, are

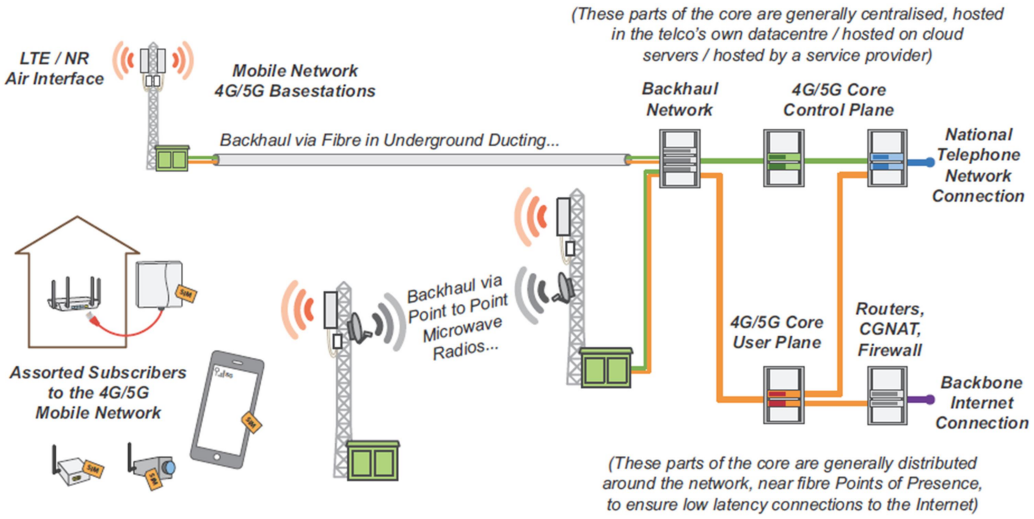


Figure 1.4: Components of a 4G/5G cellular network architecture (source: [Cro23]).

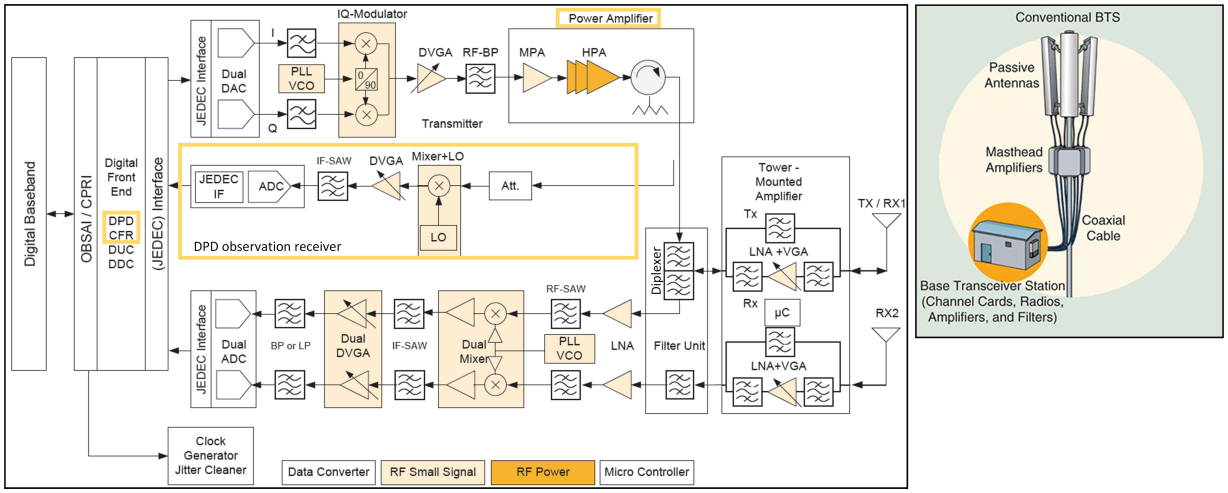


Figure 1.5: 4G FDD BTS building blocks (source: [NXP13]-left and [Wes19]-right).

interfaced to the digital front-end (DFE) which includes digital up and down converters (DUC and DDC, respectively) that perform up and down sampling and digital frequency mixing (or frequency translation) required by the digital-to-analog converters (DACs) and the analog-to-digital converters (ADCs). The DFE is also equipped with digital linearization stages like crest factor reduction (CFR) and digital predistortion (DPD) that maximize the transmitter energy efficiency while preserving linearity requirements.

The signal conversion devices interface the radiofrequency (RF) transceiver enabling proper transmission and reception over the base station antennas for a given signal bandwidth and RF band of operation in accordance to the wireless standard requirements. In the transmit path, right after the in-phase (I) and quadrature-phase (Q) DACs, these IQ signals are low-pass filtered to get rid of aliasing signals. In homodyne or zero intermediate frequency (IF) architectures, an

IQ modulator is mixing these signals with a local oscillator (LO) signal for up-conversion to RF using by using a frequency synthesizer whose main building blocks are a phase-locked loop (PLL) and a voltage controlled oscillator (VCO). It is well noted that the data conversion clocks and the PLL reference signals are typically derived from a jitter-cleaned common reference clock to have a synchronized system. Despite sharing a common clock reference, in an FDD radio the LOs will operate at a different frequency for transmission and reception (two synthesizers will be needed), while in a time-division duplexing (TDD) radio the same synthesizer can be used for the transmitter and the receiver. After the IQ modulator, a digitally-controlled variable gain amplifier (DVGA) adjusts the signal power level to compensate for the equipment ageing and manage the variation in cell load, which is followed by a RF band-pass filter (BPF) to remove out-of-band spurious signals generated by the synthesizer and the mixing process. This clean signal is delivered to the power amplification (PA) stages (i.e., a driver or medium power amplifier followed by a high power amplifier) where the signal achieves the desired level before antenna transmission. The connection between the PA and the antenna is made through a circulator (also called isolator or duplexer) in Fig. 1.5 to protect the PA against excessive power reflections and have stabilized performance under varying output port loading conditions (note that in TDD transceiver architectures the circulator is also used to isolate the transmit and receive branches). At the output of the PA stage, a signal coupler takes a portion of the transmit RF signal which is then attenuated, down-converter, power level adjusted and filtered before ADC sampling. Such receive branch conforms the so-called DPD observation receiver that is used to provide a digitized PA output baseband signal, that in conjunction with the transmit digital baseband signal is required to calculate the DPD algorithm coefficients.

In an FDD transceiver the transmitter and receiver may share simultaneously the antenna due to their operation at different RF frequencies. For such reason a diplexer, which is a three-port combiner/splitter device operating at two separated RF bands of interest, will be driving the signal to the tower-mounted amplifier that compensates for the cable losses of the long coaxial cable between the base station shelter equipment and the antennas placed in the top of the tower, and viceversa. In the receive path, if the received signal strength is high (i.e., mobile users in the vicinity of the base station) the tower mounted amplifier system will just band-pass filter the RF signal to remove potential interfering signals. When a weak signal is present at the receive antenna (i.e., distant users), the signal is not only filtered but amplified by a low-noise amplifier (LNA) and its level adjusted to meet the optimal RF receiver and ADC dynamic range. After such signal conditioning, typically a sharp response but higher-loss RF surface acoustic wave filter is put after the low noise amplification stages to remove any remaining out-of-band signal before the down conversion mixer (or IQ demodulator depending on the transceiver architecture). In the receiver of Fig. 1.5 there is a dual RF mixer with two channels down converting the signals from RF to IF, followed by an IF SAW filter, DVGA stages and IF BPFs or LPF before signal conversion. Such amount of filtering is needed in the

receive stages given the fact that a received weak signal could be in the neighbouring portion of spectrum of a transmitter close to the base station that delivers a signal strength several orders of magnitude higher at the receive antenna port. In this diagram two receivers are connected to two different antennas to benefit from spatial diversity and work with the branch having the best signal-to-noise ratio (SNR).

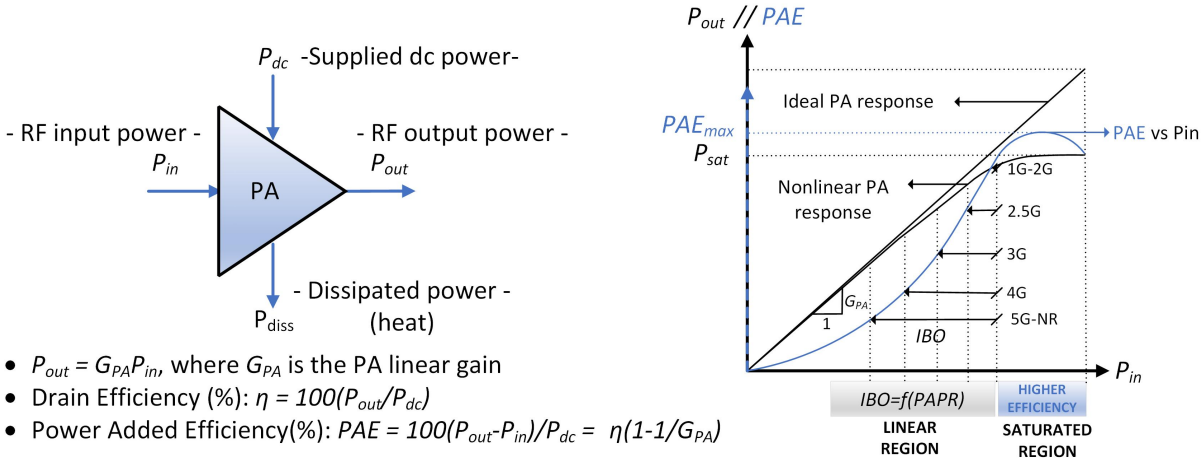
This radio architecture will not differ very significantly from that of a wireless backhauling equipment in terms of the composing building blocks. The main differences will appear in the air interface modulation, RF operating bands, signal bandwidth, power level and type of antenna, where higher-directivity dish antennas are used in wireless backhauling in contrast to the lower-directivity sector panel antennas used for radio access. The most relevant blocks that are relevant to this thesis are the digital linearization stages at the DFE, the transmitter and PA, and finally the DPD observation branch, all of which conform a closed-loop system able to adapt to changing conditions. The importance of such blocks in 4G and 5G radio access and wireless backhauling, their limitations or boundary conditions, and the most relevant technical challenges will be described in the sections following and linked to the thesis goals.

1.1.2 Need for Power Amplifier Digital Linearization

Radio Access

The beyond 4G communication systems based on orthogonal frequency division multiplexing (OFDM), and other backward compatible variants like filter-bank multi-carrier (FBMC) (a candidate 5G waveform that was not finally adopted in 5G due to its higher complexity), are based on modulation techniques which allow significantly increased spectral efficiency and capacity to mobile RANs. However, the use of these modulation techniques impacts the requirements of the radio base stations and, in particular, the PA which is a critical building block in almost all wireless communication systems. The PA is needed to strengthen the transmitted signal to compensate for the wireless radio channel losses. To do this, the PA converts a low-power input signal to a higher power one by transforming the supplied direct current (dc) power into RF power as depicted in Fig. 1.6-left. The power efficiency of the PA, however, is usually low and most of the energy is dissipated as heat thus limiting integration (i.e., requiring more complex dc supply circuits and both passive and active cooling systems) and the life-time of the devices. Non-constant envelope-modulation techniques with high peak-to-average power ratios (PAPR) require highly linear PA amplitude and phase responses to fulfill stringent spectral mask and modulation accuracy requirements to preserve the air interface performance. This is often achieved by operating the PA with significant input power level back-off (IBO) as shown in Fig. 1.6-right. This considerably reduces the PA efficiency since the PA maximum efficiency is achieved near the saturation point and only the highest waveform peaks (with very low occurrence probability) will be close

to that region.



- $P_{out} = G_{PA}P_{in}$, where G_{PA} is the PA linear gain
- Drain Efficiency (%): $\eta = 100(P_{out}/P_{dc})$
- Power Added Efficiency(%): $PAE = 100(P_{out}-P_{in})/P_{dc} = \eta(1-1/G_{PA})$

Figure 1.6: General operating principles of a PA (left) and visualization of the linearity vs efficiency trade-off (right) for different PAPR waveforms.

The PAs has been traditionally the most energy-consuming element in mobile networks and given its moderate-to-low power efficiency, most of the energy consumed is wasted as heat which in turn may require air conditioning in the equipment shelter (see Fig. 1.5) and increase the overall energy and costs. Several studies conducted during the implantation of 4G showed that the PA was typically consuming between 50% and 75% of the energy in macro-cell base stations [Deb11]. Considering also that the base stations were accounting for more than 50% of the cellular network power consumption, improving efficiency has been of high importance. The further the PAs are pushed into saturation, however, the higher the nonlinearities introducing distortion are (which in return degrades the communication quality and limit the data capacity). Therefore, different type of tools have been needed to reach the best linearity vs efficiency trade-off has been of high importance.

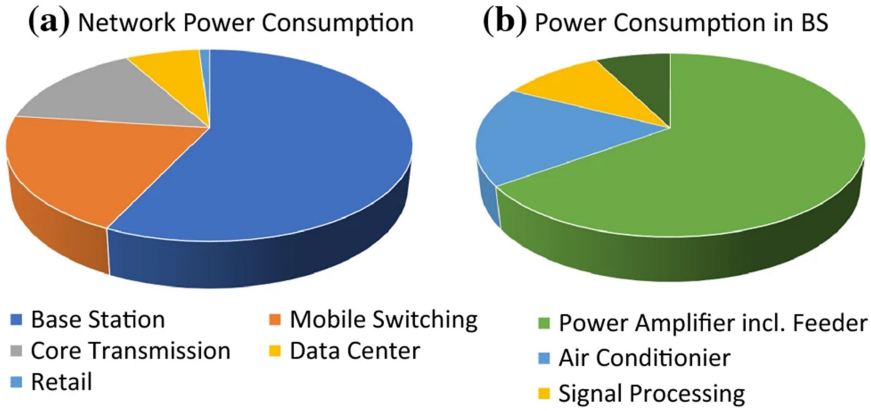


Figure 1.7: Power consumption in cellular networks (source: [Poi20]).

To overcome the previous constraints, the adoption of gallium nitride (GaN) PA technologies

and PA linearization techniques have played a key role in building better radio access and backhaul infrastructure equipment. GaN devices feature improved power-conversion efficiency, allow for higher RF power densities, higher switching speeds and bandwidth, are smaller in size than traditional devices, and relax the heat sink requirements. This translates in lower infrastructure equipment capital/operating expenses (CAPEX/OPEX) due to reduced bill of materials (BOM) and running costs which contribute in total cost of ownership (TCO) reduction. The manufacturing cost structures, however, are still too high to reach the scale volume of the silicon on insulator (SOI) or even the gallium arsenide (GaAs) industries. This has traditionally prevented the implantation of GaN devices in mobile terminals, where it is used in high-capacity and high-speed smartphone chargers, but this is something that could be changed over the next decade.

By combining such advanced PA devices, mainly in Doherty PA architectures, with digital linearization techniques like crest factor reduction (CFR) and digital predistortion (DPD), the best linearity versus power efficiency trade-off may be achieved for wireless infrastructure equipment dealing with spectrally-efficient high-PAPR waveforms. Thanks to their higher performance, flexibility and software reconfigurability (when compared to legacy analog linearization solutions) the CFR and DPD digital linearizers have become essential building blocks embedded in nowadays radio infrastructure modems.

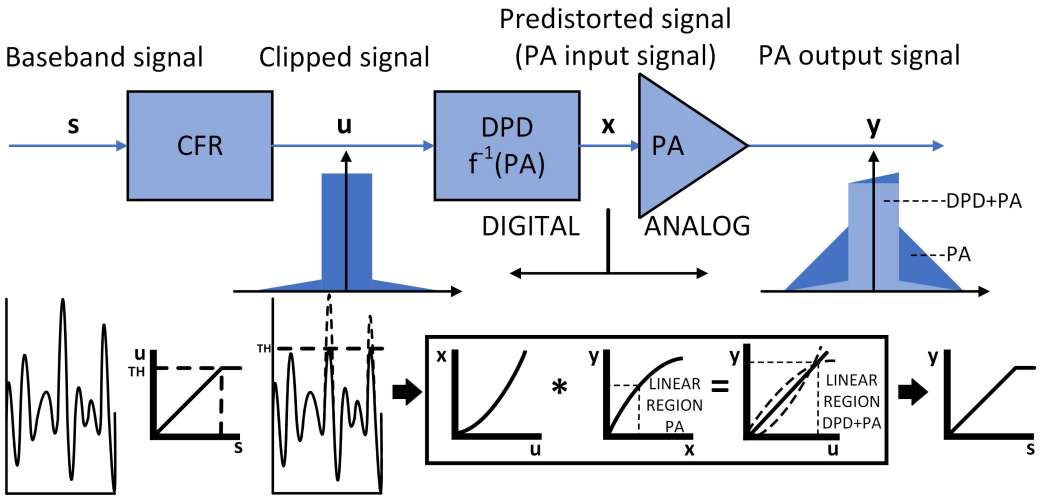


Figure 1.8: Simplified diagram showing the CFR and DPD operating principles

CFR reduces the peaks of the modulated waveform to a satisfactory level to better enable the use of the PA: due to in-band and out-of-band distortion, CFR does not introduce system gain; but, thanks to reduction in the peaks, it is possible to operate the PA at higher average power and therefore closer its saturation point, where it is most efficient as shown in Fig. 1.6. As indicated in Fig. 1.8, CFR is combined with DPD that relies on digitally cascading a nonlinear system before the PA, which provides an inverse response to the PA that provokes a linear response at

the output of the cascaded system, including the PA [Gha13]. To accomplish that, complex non-linear models are used to characterize the PA behavior taking into account bandwidth-related memory effects usually present in wideband RF PAs that may change over time and temperature. Given the need to compensate for both in-band and out-of-band distortion, the DPD digital system needs to operate at an increased sample rate and so the DAC and ADC devices. Until 4G, the transmitter has demanded real-time DPD correction and non-real-time adaptivity to changing operating conditions (i.e., temperature, ageing or even the signal configurations were traditionally changing at a lower time scale) but 5G has challenged this, as further detailed in Section 1.2.1. The operation of the CFR and DPD functions are further detailed in Chapter 2.

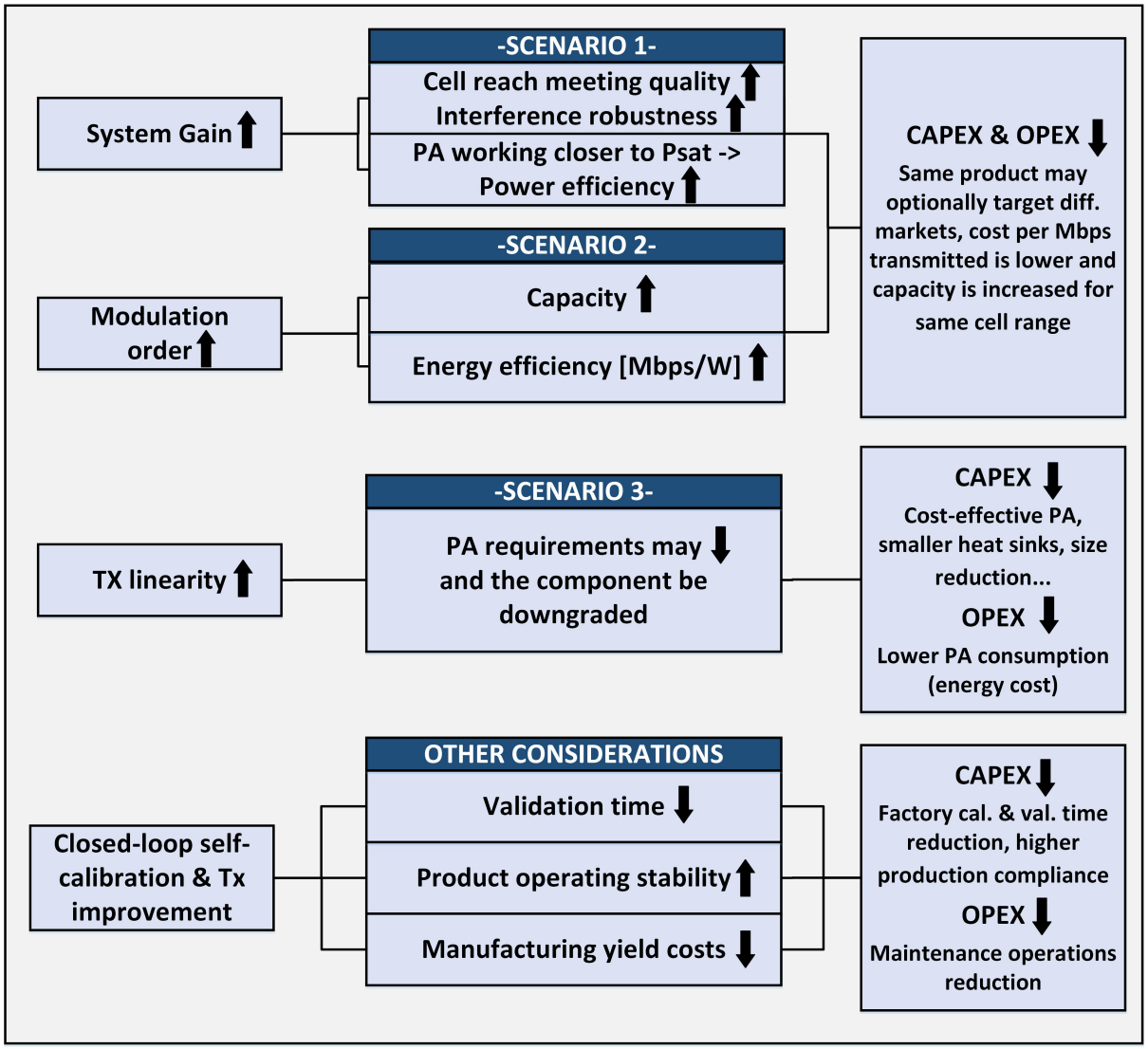


Figure 1.9: Digital linearization benefits and product level impact in radio access equipment.

As shown in Fig. 1.9, combining CFR and DPD techniques may have positive impact at product level due to the following aspects:

- The ability to combat PA nonlinearities and push forward the PA output power implies some system gain. By increasing the PA output power, the energy efficiency will be higher and the dissipated power in the form of heat will decrease. Better thermal management can also impact in the equipment reliability and prevent the active parts from degrading faster over time. Thanks to favoring the spectral containment and in-band distortion reduction, the digital linearization techniques are also an enabler not only for enhanced cell size and coverage scenarios, but new licensed/unlicensed intra-band or adjacent inter-band spectral coexistence scenarios and multi-cell deployments.

- For a given PA output power level the system performance can be enhanced therefore extending the modulation order (i.e. increasing transmission capacity and spectral efficiency) or increasing the reach and robustness against interference of radio access equipment, which maximizes the energy efficiency in terms of capacity per Watt consumed.

- Since the transmit system linearity is improved, this technique allows using cost-effective and lower power consuming PA solutions with less expensive board, enclosure and heat sink designs and simplified dc power supply and cooling subsystems.

- Closed-loop DPD may enable self-calibration and correction of the transmitter performance and therefore reduce the product validation and yield costs, increase its operating stability and minimize the number of maintenance operations.

- It is finally worth noting that the environmental impact of these solutions is twofold thanks to 1) reducing the energy consumption and CO2 emission of the RAN equipment both in production and operation, and 2) extending the product life and reducing the waste in mechanical, electrical, and electronic parts.

Wireless Backhaul

The backhaul infrastructure for mobile networks plays a role of paramount importance in the proper delivery of services requested by end users. Consequently, since the deployment of 4G the network operators have invested a significant amount of the budget devoted to network upgrades in backhaul improvement. Wireless backhauling, whose required infrastructure is significantly small, emerges as a convenient cost-effective solution, especially in countries where the deployment of wired backhauling faces economic and geographic struggles. By considering the afore introduced exponential growth of mobile broadband traffic, the cross-sector requirements aiming at increasing the overall energy efficiency of wireless networks, and the every time more important share of small cells as an alternative to handle mobile traffic, the wireless backhauling has been one of the enablers for the beyond 4G wireless communication systems, specially in those scenarios where fiber has not been previously deployed.

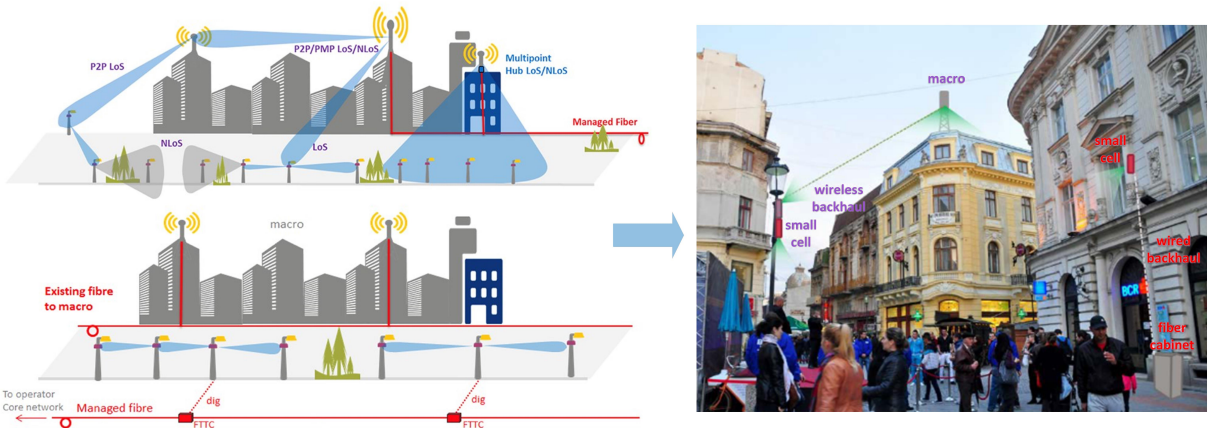


Figure 1.10: 'Macro launched' and 'street launched' small cell backhaul scenarios and deployment emulation (source: [For15]).

The traditional line-of-sight (LoS) high distance and high availability 6-42 GHz point-to-point (PtP) solutions have been kept employed for high-capacity aggregation, combined with lower distance millimeter wave (60 GHz) and E-Band (70-80 GHz) LoS PtP high-capacity solutions, and 6-GHz band LoS and NLoS PtP and PtMP medium capacity aggregation solutions.

In such a demanding scenario, reducing both CAPEX and OPEX plays a determinant role:

- Regarding the deployment costs of wireless backhaul solutions, a significant amount is due to transportation fees from the factory. Not surprisingly, the biggest part of these fees is due to the antenna size. Thus, in this scenario, reducing the size of the antennas is a key enabler for CAPEX reduction.

- Minimizing the antenna size also brings OPEX reduction since it will diminish the space occupied in the tower (and thus the rental costs) and will be less prone to pointing errors and wind misalignment. From a technology perspective, in order to reduce the size of the antennas while keeping the same link quality the output power yielded by the PA must be maximized, which implies that it must be driven as close to saturation as possible.

- Moreover, energy consumption accounts for up to 60% of the OPEX for backhauling the macro cell sites, and keeps increasing with the cost of fuel in remote sites that still rely on diesel generators for their power supply. In long haul microwave radios that are deployed on microwave towers, 90% of the power dissipation is encountered in the PA. Clearly, this puts forth reaching the best trade-off between power efficiency and linearity, where CFR and DPD play a key role.

- Increasing the output power allows operation under higher-order quadrature amplitude modulations (QAM) which increases the spectral efficiency, throughput and RAN backhauling capacity.

- The link range expansion due to having higher PA output power or the enhanced product availability, without having to increase the antenna size, is important for operation under adverse climate conditions, to deploy the backhaul system in emergency or disaster recovery situations, or to enable sub-urban/rural scenarios.

- Finally, enhancing the energy efficiency impacts in reduction of CO2 emissions as well as OPEX by simplifying enclosures, packaging and enhanced mean time between failures (MTBF).

A visual representation of the previous benefits for an E-band backhaul application scenario is provided in Fig. 1.11:

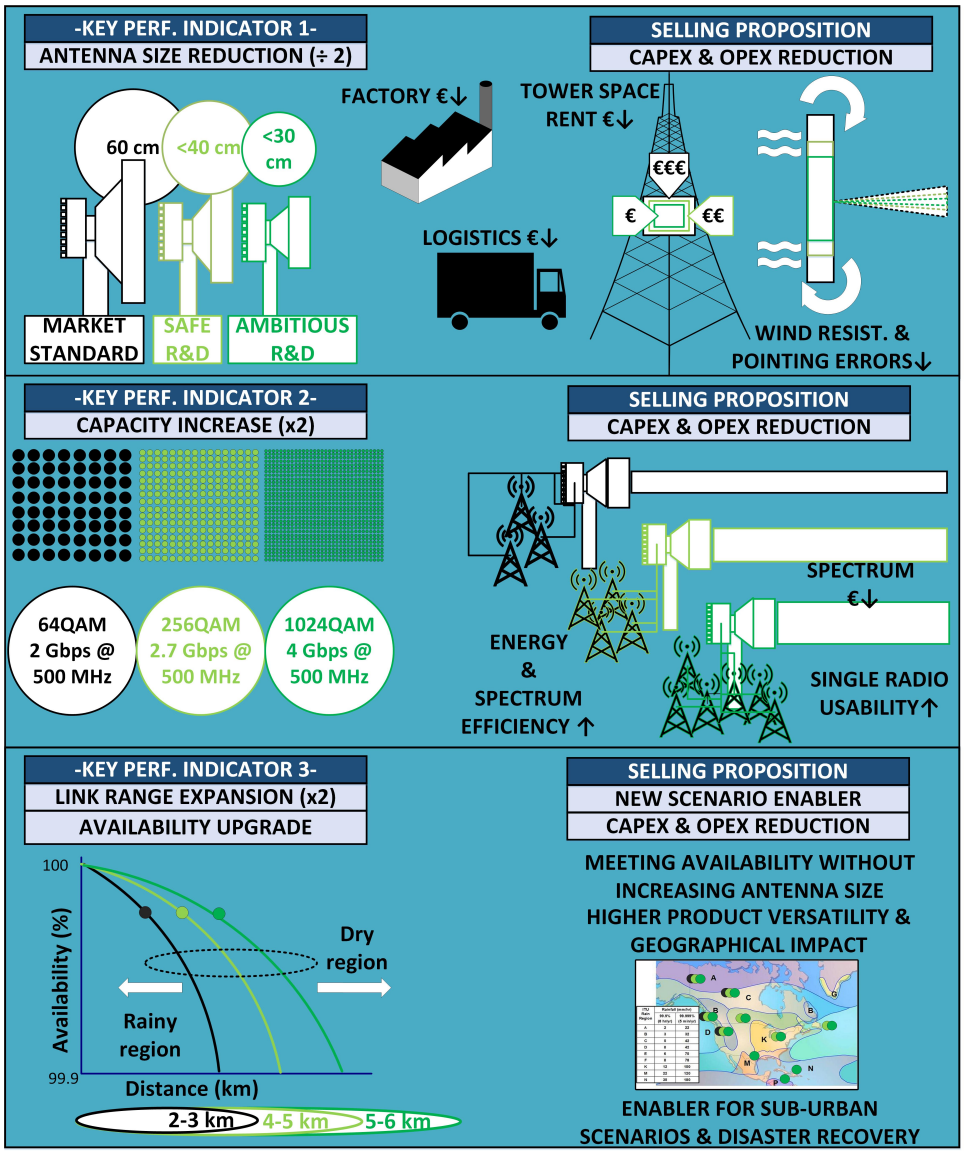


Figure 1.11: Potential benefits of combining linearization techniques in E-band backhaul radios.

1.2 Thesis Objectives and Research Contribution

1.2.1 Objectives

5G Challenges and Problem Statement

Over the last seven years the mobile broadband endless traffic increase, not only due to traditional mobile users but to the vast deployment of the Internet of Things (IoT) and machine-to-machine and human-to-machine communications (M2M and H2M), have required a more ambitious approach to target the 5G mobile networks. In order to make feasible several 5G requirements like achieving 10X increase (with respect to 4G) in terms of supported device connection density, available spectrum and peak data rates, or reduce the latency and the power consumption by 10X, clearly implies significant changes with respect to classical mobile network architectures and the physical layer technologies involved. The many technology enablers being proposed like Cloud RAN, HetNets, small-cells densification, new versatile adaptations of the physical layer schemes enabling wider bandwidths, massive multiple-input multiple-output (MIMO) transceivers and operation in millimeter wave frequencies with hybrid beamforming transceivers, deploying new licensed/unlicensed frequency bands usage paradigms, same frequency full-duplex communications, multi-hop integrated access and backhauling (IAB), device-to-device (D2D) links etc., have significantly diversified the number of different use cases and requirements.

The traditional adaptive closed-loop DPD does well in macro and micro cell deployments and typical RF and wireless backhauling radios up to 4G, where the application scenario conditions are slower varying, and the PA output power and energy consumption are significant enough to justify the increased digital signal processing complexity, DAC and ADC sampling frequency and operating bandwidth at the Tx and DPD Rx paths of the radio equipment. The vast proliferation of small cells in 5G with lower-end and lower-transmit power radios, however, makes the cost of the DPD harder to justify since the PA contribution to the base station dc power consumption breakdown becomes smaller (as depicted in Fig. 1.12) and so the potential DPD gains in terms of energy efficiency. In small cells, the contribution of the data conversion stages (ADCs and DACs) and the baseband digital signal processing (in the DFE) becomes more relevant to the overall base station dc power consumption. This makes equipping DPD in the radio system a more difficult design choice since its power cost must be reduced.

In addition, the need for new wireless communication technologies fulfilling the user mobile broadband capacity requirements, reducing costs and CO_2 emission, has boosted the research, development and industrial release of advanced MIMO transceiver solutions with a high number of elements. Thanks to focusing the energy where needed, improving spectral efficiency and capacity, and strengthening the resistance to interferers and intentional jammers, technologies

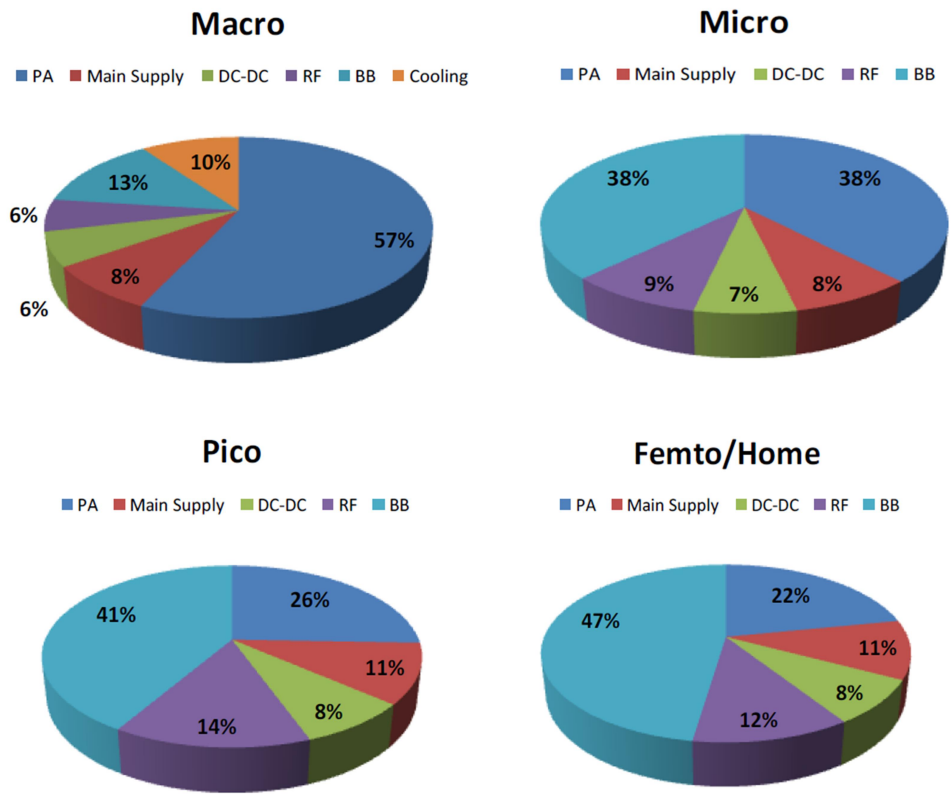


Figure 1.12: Base station power consumption breakdown for different cell sizes in 4G (source: [Aue10]).

like massive MIMO (mMIMO) are being implanted both for macro base stations and for small cells in ultra-dense deployments. The high number of RF transceiver chains in these solutions makes integration, power consumption and cost-effectiveness prominent design requirements which may play against employing the best RF performing solution. High-channel density radio frequency integrated circuits (RFICs) integrating the data conversion stages, I/Q modulators and demodulators, and LO signal synthesis and distribution are employed together with the PAs. Separate multi-antenna PA system boards are typically employed for macro and micro base stations featuring higher power, while in lower power small cell base stations (i.e., pico and femto cells) highly integrated and low-cost PAs that belong to higher-volume markets and allow integration into the RFIC transceiver are used to benefit from economies of scale. The 5G radio infrastructure PA semiconductor technology is therefore ranging from highly-performing GaN for macro cells to the cost-effective higher integration SOI complementary metal oxide semiconductor (CMOS) technologies in femtocells. The latter, however, suffers from comparably poor RF performance.

The application of DPD to 5G multi-antenna base stations is not only considered for the macro base stations but also for small cell base stations. This is due to the fact that the larger

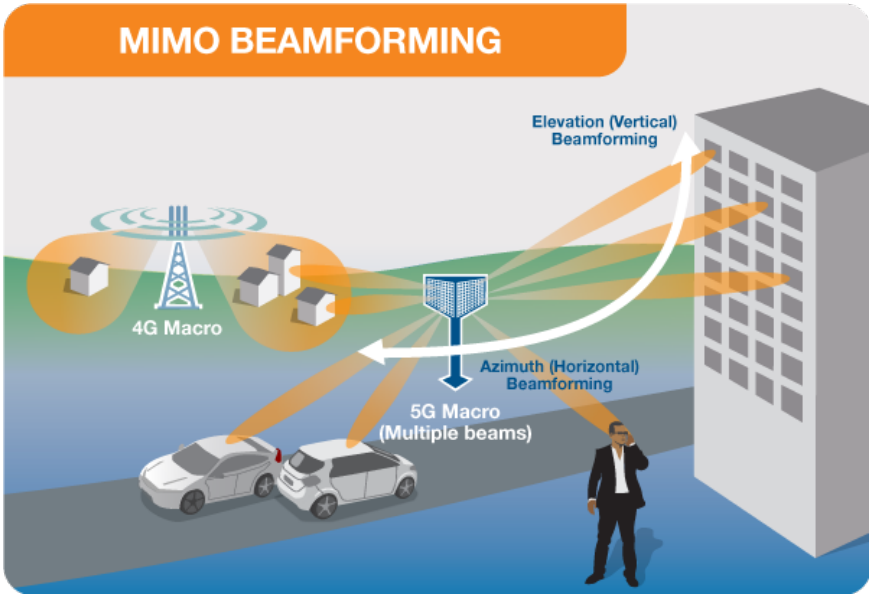


Figure 1.13: 4G sector versus 5G MIMO beamforming antennas simplified coverage concept representation (source: [MTA19]).

number of PA chains in 5G radios and the required smaller form factor and enclosures of the radio infrastructure equipment, makes thermal control a relevant issue that can contribute to justify the need for DPD despite the added complexity. In more detail, the massive MIMO radios relax the linearity profiles for each individual PA element and RF chain (since what matters is their combined action), but to be feasible the massive MIMO technology requires smaller form factor economy of scale in manufacturing. In millimeter-wave (mm-wave) small cells, deep sub-micron scale CMOS technology is vastly employed nowadays for the overall mMIMO transceiver ICs including the PAs. The use of CMOS processes brings much lower linearity when compared to other technologies like GaAs and GaN. A DPD solution with minimized resource usage may therefore increase both the RF performance and the energy efficiency of the multi-antenna small cell product that employs such cost-effective lower-end PAs by driving the PA closer to saturation which dissipates less power in the form of heat. Less heat implies that the thermal and mechanical design of the product is simplified, simpler dc power supplies can be equipped, and the reliability not only of the PA itself but other components in the radio is enhanced. This saves cost, size and weight which is a relevant requirement or enabler to 5G massive deployments. Developing lower cost energy and hardware efficient DPD systems for multi-antenna 5G systems, however, is a highly-demanding endeavor that requires several innovations going far beyond the traditional closed-loop DPD.

As shown in Fig. 1.14, the many technical challenges of 5G applications increase the complexity and power consumption of the DPD, and contravene the market need for cost-effective highly performing and fast-adapting 5G DPD solutions:

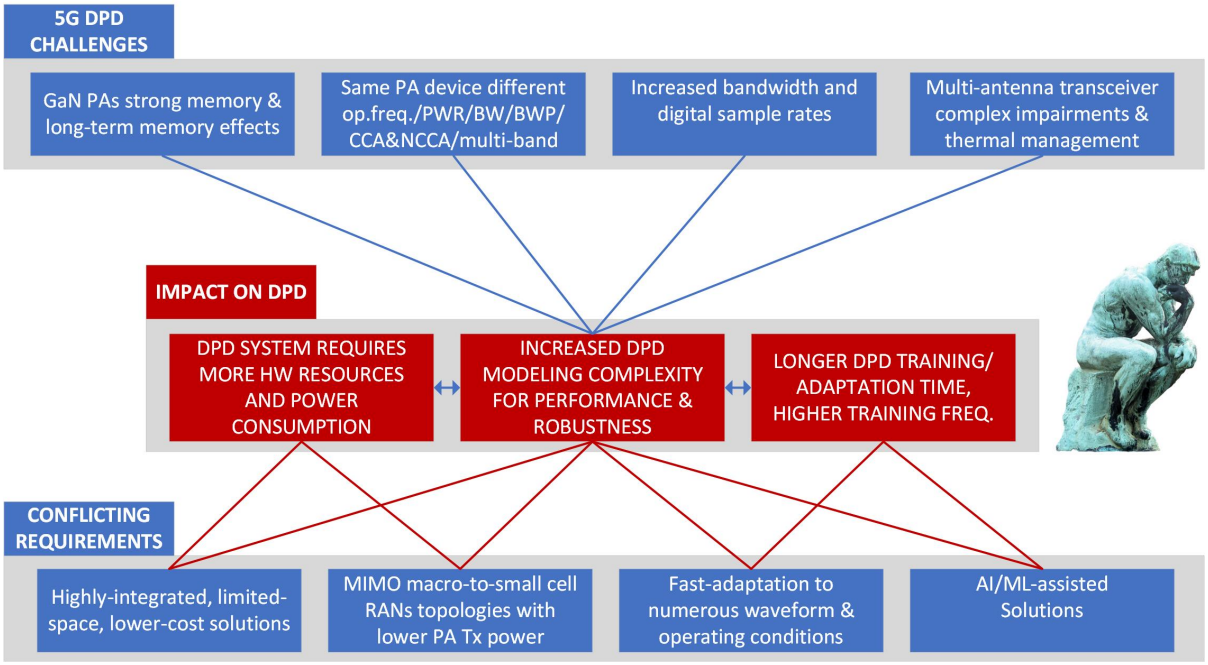


Figure 1.14: 5G DPD challenges, impact on DPD and opposite application requirements dilemma.

- Burst-like waveforms and GaN technology: While in FDD systems the downlink and uplink channel responses do not match exactly due to having different Tx and Rx frequencies, in TDD these are the same for which the MIMO/beamforming channel estimation algorithms perform better and faster. Faster channel estimation implies having more time for data transmission and therefore increased spectral efficiency. TDD operation implies also having burst-like waveforms in the transmitter, which prevents the PA noise to impact negatively on the receiver. GaN technology is mainly used in macro and micro base stations and it is foreseen that once the manufacturing costs will be downscaled it will be also vastly deployed in small cells.

  When using TDD waveforms and GaN technology, the GaN transistors shows power dependent gain variations and long-term memory effects due to trapping effects when the gate is being switched on and off [Woo17]. Such power-dependent memory effects cannot be optimally modeled with classic Volterra-based DPD models which need further adaptations to this scenario.

- Flexible, frequency-agile, fast-changing physical layer (PHY): One key feature of 5G is its PHY flexibility when compared to 4G. Despite sharing the same modulation, in 5G OFDM has been adapted to support different numerologies (i.e., subcarrier spacings) that facilitate operation with wider signal bandwidths under controlled PAPR, and minimizes the synthesizer phase noise impact in terms of intercarrier interference (ICI) when operating at higher RF frequencies. In addition, 5G signal can be set up having different bandwidth

parts, multi-band and multi-carrier configurations (i.e., with inter-band non-contiguous and intra-band contiguous and non-contiguous carrier aggregation), and be operated under digital MIMO and hybrid beamforming baseband modems that will drive uncorrelated waveforms to the PAs or the same waveform affected by different attenuation and phase variation weights, respectively.

In order to meet demanding dedicated backhauling requirements in dense deployments at street level, 5G has introduced in the RAN architecture the self-backhauled or integrated access and backhaul (IAB) radio concept. In this architecture, every IAB donor may be connected to some IAB nodes sharing either frequency, time and spatial (beam) resources, where the IAB node is a combination of a gNB (BS) and a UE and thus plays the role of access node and backhaul relay node simultaneously. Operating efficiently the PAs in such a versatile RAN architecture requires complex adaptive DPD models increasing the hardware (HW) complexity to meet the required quality performance, robustness and minimized radio interference. Short DPD training periods are also needed to support the fast adaptivity requirement (i.e, up to millisecond-range level), which is in conflict with having to train complex DPD models.

- Wider bandwidths: 5G has to support much wider signal bandwidths when compared to 4G. While up to 100 MHz bandwidth signals are managed in 5G NR frequency range 1 (FR1) sub-7 GHz radios, in FR2 mmwave Ka-bands the signal bandwidth per channel may be of up to 400 MHz. When the radio is designed to support carrier-aggregation, the intra-band contiguous and non-contiguous bandwidths may scale up to 400 MHz in FR1 and 1200 MHz in FR2. This challenges the DPD with respect to the following aspects: The higher the signal bandwidth, the higher the memory effects of the PA and the more complex the modeling techniques needed be to characterize and compensate for the PA nonlinearities. In order to compensate for the out-of-band distortion in DPD, such wide signals need to be oversampled to cover sufficient out-of-band compensation, since not only the in-band error-vector magnitude (EVM) but also the adjacent channel power ratio (ACPR) requirements must be fulfilled.

  Operating the DFE logic and the data conversion circuits at such high speeds, not only increases the digital implementation complexity and timing constraints and jeopardizes cost-effectiveness, but the energy consumption of the DPD system building blocks, all of which makes harder to justify the DPD. Moreover, when increasing the bandwidth of the signal being observed in the DPD feedback path, the received noise will be higher and will degrade the quality in DPD modeling unless averaging is applied to mitigate the SNR reduction.

- Complex nonlinear effects multi-antenna transmitters: In 5G, both when operating with digital MIMO (FR1) or hybrid beamforming (FR2) transceiver architectures, the PAs

will not be receiving the same signal. In addition, despite the discrete PAs in a multi-antenna system will share the same characteristics and manufacturing processes, they will not behave exactly the same. If there are small differences, these will be aggravated with temperature and supply variations or due to aging. The straight-forward approach to design a high-end multi-antenna DPD, for a macro BS or RRH, is therefore to consider one DPD system per PA branch. However, for large systems this is not economically viable and the complexity and power consumption of such approach will be unaffordable.

When considering lower-power highly integrated approaches in multi-antenna small cells having RFICs having numerous PA branches (and less manufacturing tolerances), there are several impairments that can lead the previous approach to failure. This is due to the cross-effects that may appear in these architectures, mainly PA input and output cross couplings and PA loading effects in circulator-less architectures (i.e., in TDD). This scales very significantly the dimensionality of the behavioral models that need to account for such interactions, and can boost dramatically the DPD implementation complexity even for a reduced number of antennas. Finally, while digital linearization techniques can help to control thermal issues due to power dissipated as heat, the increased sample frequency requirements at the data conversion devices can also provoke thermal issues when having to consider compact format radios.

The previous challenges that have been presented mainly for 5G radio access are also applicable to pure wireless backhauling systems. Given the fact that the emerging network deployments demand higher capacity, spectral efficiency/containment/flexibility, energy efficiency and cost reduction, the operation at mm-wave frequencies is an enabler to provide 5G and B5G cellular and backhaul communications thanks to its massive unlicensed and lightly licensed bandwidth. To meet these requirements in the challenging mm-wave propagation environment poses several threats which may impair the quality of service (QoS) in vertical applications like the high pathloss and materials attenuation, or the limited transceiver performance (both in terms of noise and linearity). At these frequencies, the most important limiting factors are the PA's linearity and power efficiency. These aspects become more critical when complex multi-level modulation schemes presenting high PAPR are used to increase the bit rate without increasing the bandwidth, such as 1024-QAM at E-band and up to 4096-QAM at lower frequency bands.

The mm-wave PAs must operate well below its full capacity, which results to poor power efficiency (i.e., 5-10% values) and lower maximum output power and shorter radio hops or cell coverage. Consequently, the radio equipment and components will have to innovatively address the trade-off between bit rate (and overall linearity performance) and power efficiency (extremely low at mm-wave frequencies) by enhancing the energy per bit efficiency with higher linear output power and lower power consumption.

Research Objectives

As previously detailed, the traditional DPD approach is no longer valid to handle a significant number of 5G-like scenarios and higher complexity DPD systems are required to operate the proposed radio architectures. The digital processing resources devoted to DPD need to fit efficiently the current field-programmable gate array (FPGA) logic clocking speed and capacity. Very large or complex DPD algorithms with expanded HW requirements that go beyond using any current cost-effective digital processing technology, or requiring training times going far beyond the millisecond order, are unlikely to be widely adopted. To combat these threats, the objectives of the thesis are those summarized in Fig. 1.15:

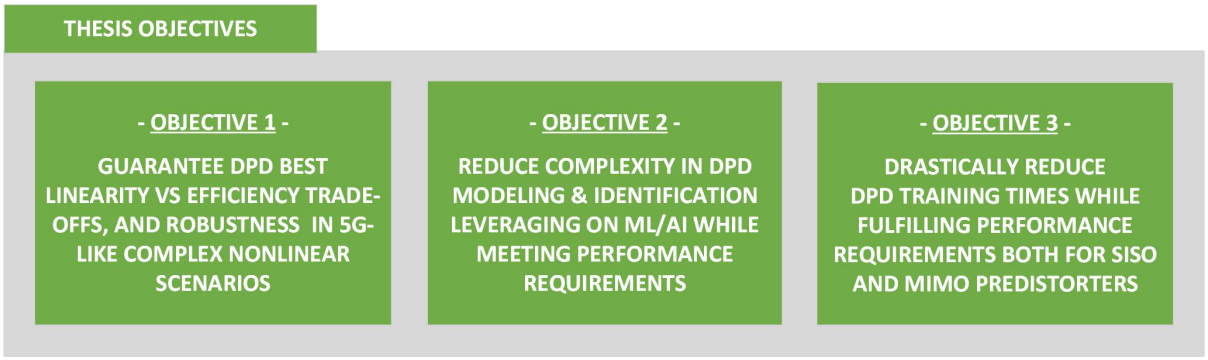


Figure 1.15: Thesis objectives addressing 5G DPD needs.

- OBJECTIVE 1: Propose DPD modeling techniques able to handle the GaN PA technology strong nonlinearities, trapping and long-term memory effects for wideband high-PAPR burst-like waveforms, and provide sufficient statistical representativity and robustness, that combined with CFR reach the best linearity versus energy-efficiency trade-offs and also minimize interference between homogeneous and heterogeneous wireless communication systems. Demonstrate the benefits in terms of standard quality and efficiency parameters like the normalized mean square error (NMSE) between the nominal baseband Tx signals and those at the PA outputs, EVM, ACPR, drain efficiency, robustness against adverse effects, or better/increased/ubiquitous coverage scenarios with higher signal to interference and noise ratio (SINR), potentially enabling new licensed/unlicensed intra-band or adjacent inter-band spectral coexistence scenarios.

- OBJECTIVE 2: Reduce the DPD hardware implementation complexity by using model order reduction techniques, based on ML feature selection and feature extraction methods, to reduce the number of DPD coefficients in the forward application DPD Tx path and/or the feedback Rx DPD path where model parameters or DPD coefficients are estimated to minimize the error between the nominal DPD oversampled Tx baseband signal

and the one obtained at the PA output (i.e., after down conversion, gain correction and baseband time alignment). These techniques not only have to meet complexity reduction but guarantee well-conditioned DPD estimation and meet the radio performance requirements. The current communication network technology trends aiming to deploy machine learning (ML) and artificial intelligence (AI) in the RAN, not only to replace physical layer building blocks but to manage the RAN radio and computational resources. In line with this promising trend, the hardware-efficient DPD systems need to be supported by ML techniques and be leveraged both for polynomial and artificial neural network (ANN) DPD architectures, both for SISO and (more importantly) MIMO transceiver architectures.

- OBJECTIVE 3: Combine novel training data selection mechanisms with the proposed ML-assisted DPD model order reduction techniques to dramatically reduce the polynomial DPD and ANN DPD training times (extremely large in ANNs) for complex systems, and to target 5G's fast-adaptivity and performance requirements both for SISO and MIMO DPD architectures.

Therefore, the thesis aims to research, design, combine and experimentally evaluate DPD techniques that meet the previous objectives. While the thesis experimental validation of the novel DPD algorithms will be conducted considering sub-7 GHz radio equipment, the proposed techniques have also potential use in mm-wave radios since they have been applied to DPD linearizers that operate at baseband and are independent on the final RF frequency band.

1.2.2 Contribution and Novelty

Prospective Works

The prospective works are those that leverage on previously existing techniques, or convenient fine-grained adaptations, that are properly combined to address the needs of a specific challenging linearization scenario. At the end of every work description, both the targeted thesis objectives and the publications that have been generated are mentioned.

Strategies to combat strong nonlinearities and long-term memory effects in GaN PAs under burst-like waveforms: This contribution combines several strategies to deal with GaN PAs power dependent gain variations and long-term memory effects due to trapping effects, when burst-like waveforms that switch on and off the transistors gate are employed for 4G/5G. For a given amplitude threshold, the solution proposed first includes already existing decomposed piecewise curve-fitting DPD model considering the memory polynomial (MP) model for the samples equal or above the threshold and the generalized memory polynomial (GMP) model for the samples below the threshold. Second, an additional GMP model with no amplitude discrimination is then multiplied by a finite sample size sliding average input power window to

handle long-term memory effects. The low amplitude, high amplitude and long-term memory modeling structures are then extended in number of observations after application of several training data waveforms and third, a two-dimensional mesh-selecting mechanism is applied to these data batches to reduce the number of observations by selecting the most statistically representative ones. Finally, before the DPD system a peak-cancelation CFR technique is also applied to partially get rid of low-probability waveform peaks and allow trading-off between DPD complexity and performance.

Target: OBJECTIVE 1. Publications: Magazine article (P3) and conference paper (P4).

Evaluation of CFR and DPD for FBMC waveforms in spectrum-shared access: For the first time, the use of digital linearization techniques in a radio DFE is evaluated not only in terms of linearity and efficiency gains, but spectral coexistence between heterogeneous systems, in terms of signal-to-interference-plus-noise ratio (SINR). This work is positioned in the context of a broadband public protection and disaster relief (BB-PPDR) radio communication scenario, where a spectrum sharing cognitive filter-bank multicarrier (FBMC) system is used as private mobile radio (PMR) system that coexists with other legacy systems such as 4G in an emergency deployment. The need for frequency-agile spectrally-contained waveforms, high spectral-efficiency and data capacity in an energy-constrained scenario justifies the use of FBMC together with CFR and adaptive DPD.

Target: OBJECTIVE 1. Publication: Book chapter (P9).

Feature selection and feature extraction early works: With regard to feature selection techniques, a new variant of the orthogonal matching pursuit (OMP) algorithm for reducing the computational complexity of the DPD behavioral model in the forward path is presented. The proposed spectral weighting OMP (SW-OMP) algorithm focuses on selecting the most relevant basis functions to compensate for the out-of-band residual distortion which may eventually be masked by the dominant in-band residual error. This basis selection is carried out in an off-line process that does not affect the computational complexity of the real-time closed-loop DPD but, on the contrary, reduces its complexity while enhancing the robustness. Experimental results with 4G radio access intra-band non-contiguous carrier-aggregated waveforms show that by selecting the DPD coefficients with SW-OMP, the inherent ACPR and NMSE degradation suffered when reducing the number of coefficients is mitigated under strong nonlinear operation, when compared to using the basis functions selected by the classical OMP algorithm. SW-OMP is therefore a more appropriate dimensionality reduction technique for spectrum sharing scenarios. With respect to feature extraction techniques being applied to the DPD feedback path, the PC-CFR scheme is combined with a principal components analysis (PCA) pruned DPD algorithm. This approach allows to enhance the transmitted output power of a microwave backhaul transmitter while reducing the complexity of the DPD estimation process, and meet the performance required by the application.

Target: OBJECTIVE 1 and OBJECTIVE 2. Publications: Conference papers (P1) and (P2), and journal article (P6) (co-author).

Main Research Contributions

The main research contributions of the thesis consist in either coarse-grained adaptations of existing algorithms, their novel application to solve challenges in complex linerization scenarios, and newly proposed techniques to overcome limitations with respect to the SotA. These contributions, which are further detailed in Section 3.2, Section 4.2 and in Chapter 5, are summarized as follows:

- Independent DPD parameters estimation method for SISO adaptive polynomial DPD: In order to optimize the resources employed for DPD estimation and reduce training time, an estimation/adaptation method based on the (feature extraction) adaptive principal components analysis (APCA), applied considering a convenient modification of the complex domain generalized Hebbian algorithm (CGHA), is presented to guarantee the identification of the minimum necessary parameters of a digital predistorter. The proposed estimation/adaptation technique is suitable for online FPGA or system on chip (SoC) implementation. By exploiting the orthogonality of the resulting transformed matrix obtained with the APCA technique, it is possible to reduce the number of coefficients to be estimated which, at the same time, has a beneficial regularization effect by preventing ill-conditioning or over-fitting problems. Therefore, this identification/adaptation method enhances the robustness of the parameter estimation and simplifies the adaptation by reducing the number of estimated coefficients. Due to the orthogonality of the new basis, these parameters can be estimated independently, thus allowing for scalability which allows to reach controlled trade-off the DPD adaptation time versus performance and hardware complexity.

  Target: OBJECTIVE 1 and OBJECTIVE 2. Publication: Journal article (P5).

- Training data selection and model order reduction techniques for SISO adaptive ANN DPD: Despite the good approximation capabilities of ANNs for DPD modeling and their enhanced performance over polynomial modeling, these typically have higher number of DPD coefficients (i.e., weights and biases) and require significantly larger DPD training times that cannot be supported in fast-changing 5G applications. To overcome such limitations, the already existing doubly orthogonal matching pursuit (DOMP) algorithm is applied for the first time to reduce the number of input dataset features or variables which reduces the weights in the first hidden layer. The consecutive and sparse batch selection mechanisms (CBS, SBS) are proposed to select training waveform data batches according

to performance metrics that can be adapted for a given scenario or targeted key performance indicators, outperforming the existing multi-dimensional histogram-based techniques in reduction factor and significantly reducing the DPD updating time.

Target: OBJECTIVE 1, OBJECTIVE 2 and OBJECTIVE 3. Publications: Conference papers (P7) and (P10), and book chapter (P8) (co-author).

- Training data selection and model order reduction techniques for MIMO adaptive polynomial and ANN DPD: In MIMO DPD architectures, given the large number of antennas and having to consider cross-effects, both the polynomial behavioral modeling matrices and MIMO ANN datasets may have unaffordable size leading to ill-conditioning, performance instabilities and long training time periods that need to be drastically reduced for challenging 5G and beyond scenarios. The combination of an efficient uncorrelated equation selection (UES) mechanism together with orthogonal least squares (OLS) is proposed to reduce the training data length and the number of basis functions, respectively, at every behavioral modeling matrix in the polynomial MIMO DPD scheme. For ANN MIMO DPD architectures, applying UES and principal component analysis (PCA) is proposed to reduce the input dataset length and features.

  Combining training data selection methods and dimensionality reduction techniques was not previously addressed in literature for MIMO transmitters with strong PA input and output cross talk provoking complex nonlinear effects. The existing SotA histogram-based sample or equation selection methods to select training data samples do not operate well under such conditions and run slower than the proposed UES mechanism. With regards to the art in dimensionality reduction techniques, the novel application of OLS and PCA to MIMO DPD is formulated both for polynomial modeling basis or ANN dataset feature selection. While in polynomial MIMO DPD, OLS is the best option, in ANN MIMO DPD, PCA and OLS can be used to trade-off between ANN MIMO DPD performance and reduced complexity (and faster adaptivity). The procedure to apply the methods is fully detailed for one polynomial and two ANN MIMO-DPD architectures and their goodness is experimentally validated against the SotA reference methods thus demonstrating the benefits in terms of complexity and DPD update/training time reduction while reasonably meeting the performance requirements at the base station side.

  Target: OBJECTIVE 1, OBJECTIVE 2 and OBJECTIVE 3. Publication: Journal article (P11).

Publications resulting from the thesis research

The research reported in this thesis has generated publications in international and national conferences, international indexed scientific journals, and book chapters. These publications are listed in chronological order in the following.

(P1) **D. López-Bueno**, P. L. Gilabert, G. Montoro, N. Bartzoudis, "Peak cancellation and digital predistortion of high-order QAM wideband signals for next generation wireless backhaul equipment," *in Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-w. Circ.*, 2-4 April 2014, Leuven (Belgium).

(P2) **D. López-Bueno**, P. L. Gilabert, G. Montoro, N. Bartzoudis, M. Payaro, "Reducción de Factor de Cresta y Predistorsión Digital de Señales QAM de Banda Ancha para Equipos Backhaul de Próxima Generación," *in Proc. XXX Simp. Nac. Unión Científ. Int. Radio*, 2-4 September 2015, Pamplona (Spain).

(P3) **D. López-Bueno**, T. Wang, P. L. Gilabert, G. Montoro, "Amping Up, Saving Power: Digital Predistortion Linearization Strategies for Power Amplifiers Under Wideband 4G/5G Burst-Like Waveform Operation," *IEEE Microw. Mag.*, Vol. 17, No. 1, pp. 79 - 87, January 2016 (ranked **Q1** 2016 by JCR).

(P4) **D. López-Bueno**, P. L. Gilabert, G. Montoro, N. Bartzoudis, O. Font-Bach, "SHAPER: Demostrador Experimental de Técnicas de Linealización Digital de Amplificadores de Potencia para Acceso y Backhaul Inalámbrico 4G/5G,"*in Proc. XXXI Simp. Nac. Unión Científ. Int. Radio*, 5-7 September 2016, Madrid (Spain).

(P5) **D. López-Bueno**, Q. A. Pham, G. Montoro and P. L. Gilabert, "Independent Digital Predistortion Parameters Estimation Using Adaptive Principal Component Analysis," *in IEEE Trans. Microw. Theory and Tech.*, vol. 66, no. 12, pp. 5771-5779 (ranked **Q1** 2018 by JCR).

(P6) P. L. Gilabert, **D. López-Bueno**, T. Wang, G. Montoro, "Spectral Weighting Orthogonal Matching Pursuit Algorithm for Enhanced Out-of-Band Digital Predistortion Linearization," *in IEEE Trans. Circ. Systems II: Expr. Briefs*, vol. 66, no. 7, pp. 1277-1281, July 2019 (ranked **Q2** 2019 by JCR).

(P7) **D. López-Bueno**, Q. A. Pham, G. Montoro and P. L. Gilabert, "Linealización digital de transmisores mediante redes neuronales no lineales," *in Proc. XXXIV Simp. Nac. Unión Científ. Int. Radio*, 5-7 September 2019, Sevilla (Spain).

(P8) P. L. Gilabert, **D. López-Bueno**, T. Quynh Anh Pham, and G. Montoro, "Machine learning for digital front-end," *in Machine Learning for Future Wireless Communications.* John Wiley & Sons, Ltd, 2020, ch. 17, pp. 327-381.

(P9) **D. López-Bueno**, N. Bartzoudis, O. Font-Bach, M. Caus, P. L. Gilabert, G. Montoro, "Technologies for emergency rollout of broadband public protection and disaster relief (BB-PPDR) communications in humanitarian crisis zones," *Chapter in Information and communication technologies for humanitarian services*, published by The Institution of Engineering and Technology, August 2020, pp. 215-246. ISBN 978-1-78561-996-0.

(P10) **D. López-Bueno**, P. L. Gilabert and G. Montoro, "Dataset Reduction for Neural Network Based Digital Predistorters under Strong Nonlinearities," *IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, 2021, pp. 8-11.

(P11) **D. López-Bueno**, G. Montoro, and P. L. Gilabert, "Training Data Selection and Dimensionality Reduction for Polynomial and Artificial Neural Network MIMO Adaptive Digital Predistortion," *IEEE Trans. Microw. Theory Techn. 2022 AI Spec. Issue*, doi: 10.1109/TMTT.2022.3209214 (ranked in top **Q2** 2022 by JCR).

Other Publications on DPD as Co-Author

In the context of the thesis topics and R&D activities, the author has also contributed to the following publications carried out by the Research Group:

(P12) P. L. Gilabert, G. Montoro, **D. López-Bueno**, N. Bartzoudis, E. Bertran, M. Payaro, A. Hourtane, "Order reduction of wideband digital predistorters using principal component analysis," *in Proc. IEEE MTT-S Int. Microw. Symp.*, 2-7 June 2013, Seattle, WA (USA).

(P13) P. L. Gilabert, G. Montoro, **D. López-Bueno**, N. Bartzoudis, M. Payaro, A. Hourtane, "Reducción de Orden en Predistorsión Digital para Aplicaciones de Banda Ancha," *in Proc. XXVIII Simp. Nac. Unión Científ. Int. Radio*, 2-4 September 2013, Santiago de Compostela (Spain).

(P14) G. Montoro, T. Wang, **D. López-Bueno**, M. N. Ruiz, J. A. García, P. L. Gilabert, "Reducción de la Frecuencia de Muestreo en los Conversores ADC y DAC usados en Predistorsionadores Digitales," *in Proc. XXX Simp. Nac. Unión Científ. Int. Radio*, 2-4 September 2015, Pamplona (Spain).

(P15) Q. A. Pham, **D. López-Bueno**, G. Montoro, and P. L. Gilabert, "Adaptive principal component analysis for online reduced order parameter extraction in PA behavioral modeling and DPD linearization," *in Proc. IEEE MTT-S Int. Microw. Symp.*, June 2018, pp. 1-4.

(P16) Q. A. Pham, **D. López-Bueno**, T. Wang, G. Montoro, and P. L. Gilabert, "Multidimensional LUT-based digital predistorter for concurrent dual-band envelope tracking

power amplifier linearization," *IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, Jan. 2018, pp. 1-4.

(P17) Q. A. Pham, **D. López-Bueno**, T. Wang, G. Montoro and P. L. Gilabert, "Partial Least Squares Identification of Multi Look-Up Table Digital Predistorters for Concurrent Dual-Band Envelope Tracking Power Amplifiers," *in IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5143-5150, Dec. 2018 (ranked **Q1** 2018 by JCR).

(P18) Q. A. Pham, **D. López-Bueno**, G. Montoro and P. L. Gilabert, "Dynamic Selection and Update of Digital Predistorter Coefficients for Power Amplifier Linearization," *IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, 20-23 Dec. 2019, pp. 1-4

(P19) Q. A. Pham, G. Montoro, **D. López-Bueno** and P. L. Gilabert, "A method for selecting online the coefficients to be updated in a DPD for PA linearization," *in Proc. XXXIV Simp. Nac. Unión Científ. Int. Radio*, 5-7 September 2019, Sevilla (Spain).

(P20) Q. A. Pham, G. Montoro, **D. López-Bueno** and P. L. Gilabert, "Dynamic Selection and Estimation of the Digital Predistorter Parameters for Power Amplifier Linearization," *in IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 3996-4004, Oct. 2019 (ranked top **Q2** 2019 by JCR).

(P21) W. Li, N. Bartzoudis, J. R. Fernández, **D. López-Bueno**, G. Montoro and P. Gilabert, "Slow-Envelope Shaping Function FPGA Implementation for 5G NR Envelope Tracking PA," *in Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-w. Circ.*, 2022, pp. 1-3

(P22) W. Li, N. Bartzoudis, J. R. Fernández, **D. López-Bueno**, G. Montoro and P. Gilabert, "FPGA Implementation of a Linearization System for Wideband Envelope Tracking Power Amplifiers," *IEEE Trans. Microw. Theory Techn. INMMIC Spec. Issue*, doi: 10.1109/TMTT.2022.3217842.(ranked top **Q2** 2022 by JCR).

1.2.3 Other merits

Awards and Participation in Scientific Society Activities

(M1) Co-recipient of the First prize in the "Power Amplifier Digital Pre-distortion Linearization Design Competition" at the International Microwave Symposium in Phoenix (2015):

(a) (b)

Figure 1.16: (a) IMS15 Student Design Competition and (b) 1st prize award (source: LylePhotos.com, LLC).

The author teamed up with Teng Wang, also pursuing a PhD at Universitat Politècnica de Catalunya (UPC), to obtain the highest score in the Power Amplifier Linearization Through Digital Pre-distortion Linearization international competition. The event took place in the frame of the Student Design Competitions organized by the IEEE Microwave Theory and Techniques Society (MTT-S) during the International Microwave Symposium (IMS) held in Phoenix (AZ, USA) on 17-22 May 2015. The winners at every edition of the competition can be found online at dpdcompetition.com/sdc/winners/.

The author would like to thank Prof. Christian Fager, Prof. Thomas Eriksson, Dr. Koen Buisman and Mr. Sebastian Gustafsson from Chalmers University of Technology, Dr. Per Landin from Ericsson and Dr. Takao Inoue from National Instruments for their effort in preparing the competition and building the Weblab, a wonderful initiative fostering research in digital linearization and enabling worldwide access to state-of-the-art instrumentation.

(M2) Co-organizer of the "Power Amplifier Digital Pre-distortion Linearization Design Competition" at the International Microwave Symposium in Hawaii (2017):

(a) (b)

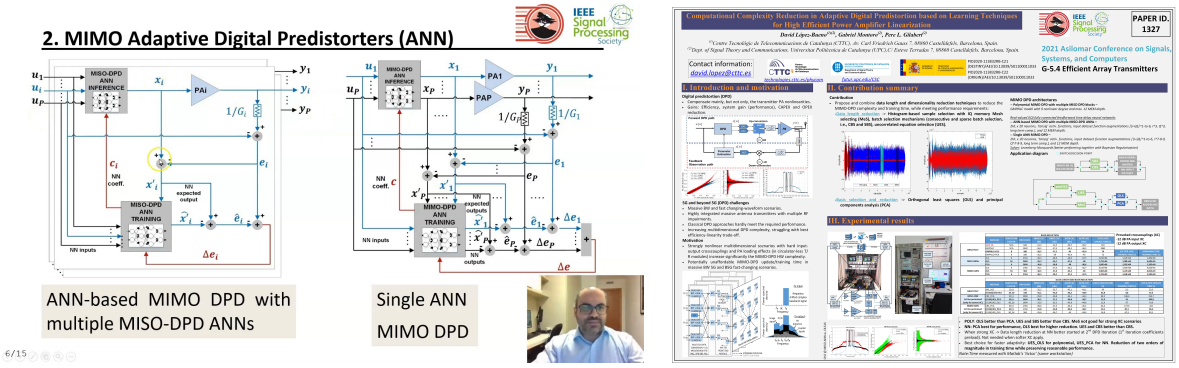
Figure 1.17: (a) IMS17 Student Design Competition (source: LylePhotos.com, LLC) and (b) final ranking with all competing teams score.

Dr. Pere L. Gilabert and Dr. Gabriel Montoro, professors at Universitat Politècnica de Catalunya (UPC), together with the author, organized the 2017 edition of the Power Amplifier Linearization through Digital Pre-distortion Student Design Competition (SDC#6), sponsored by Rohde & Schwarz and the IEEE MTT-S Digital Signal Processing (MTT-9) and Microwave Measurements (MTT-11) Technical Committees. The competition was celebrated during the International Microwave Symposium (IMS) by 4-9 June 2017 in Honolulu (HI, USA). The participants were able to access via world wide web (www) the remoteUPCLab to interface the test and measurement equipment. For the first time in the history of the DPD competition, in this edition the participants were able to participate in two DPD competitions. Prior to the IMS dates, an online competition was held. On that year's edition, record participation was achieved with eight student teams participating both in the online competition and in the IMS 2017 competition celebrated in Honolulu. Students and researchers of 27 countries were users of the remoteUPCLAB during the period that was made available.

The author would like to thank the support from Rhode & Schwarz and Mr. Hermann Boss in building the ET-DPD test setup and sponsoring the competitions.

(M3) The author also participated in an Invited talk and Poster session at the G-5 Energy-Efficient Array Transmitters Session of the 2021 Asilomar Conference on Signals, Systems, and Computers celebrated in Pacific Grove (CA, USA):

- **D. López-Bueno**, W. Li, P. L. Gilabert and G. Montoro, D. "Computational Complexity Reduction in Adaptive Digital Predistortion based on Learning Techniques for High Efficient Power Amplifier Linearization,": (Invited Talk)," 2021 55th Asilomar Conference on Signals, Systems, and Computers, 2021



(a) (b)

Figure 1.18: ASILOMAR 2021 (a) Invited talk presentation (hybrid format) and (b) poster.

Student Supervision

(M4) Three students have been supervised or co-supervised in the field of digital linearization:

- G. Pojani, "FPGA implementation of a digital pre-distorter for wideband microwave backhaul systems," Master Thesis (UNIBO-UPC), Supervisors: N. Bartzoudis (CTTC), P. L. Gilabert (UPC), **D. López-Bueno** (CTTC). Advisor: R. Verdone (UNIBO), October 2014.
- A. Mars and M. Cañal, "Evaluation of Digital Linearization Techniques for PAs under 5G Waveform Operation in Spectrum Sharing Scenarios," Bachelor Thesis (UPC-EETAC), Supervisor: **D. López-Bueno** (CTTC), Advisor: P. L. Gilabert (UPC), September 2016.

Creation of the SHAPER modular digital linearization test bed

The strong applied component of this thesis made creating a test bed a necessary outcome to provide both a simulation framework and a measurement infrastructure to program and evaluate in end-application realistic conditions the performance of the novel algorithms being proposed. The SHAPER (Solutions for wideband Highly lineAr and efficient PowER amplification) test bed has therefore been used to build both the simulation and hardware test benches and thus validate the research conducted in this thesis in a relevant environment.

SHAPER [LB22a] is a Software Defined HW demonstrator that provides rapid and cost-effective prototyping and validation of RF and uW PA linearization techniques, such as DPD and CFR. The demonstrator nowadays comprises several MATLAB test benches including the baseband Tx PHY, Rx PHY, DPD & CFR blocks, which are able to interface both high-end laboratory instruments for waveform generation and capture or performance evaluation, and COTS boards (such as FPGA-based waveform capture and data generation, ADC and DAC evaluation boards), as seen in Fig. 1.19. SHAPER can also enable HW/software co-simulation and validation of DPD and CFR building block implementations in FPGA bridging the gap between applied research and end-product prototyping and validation.

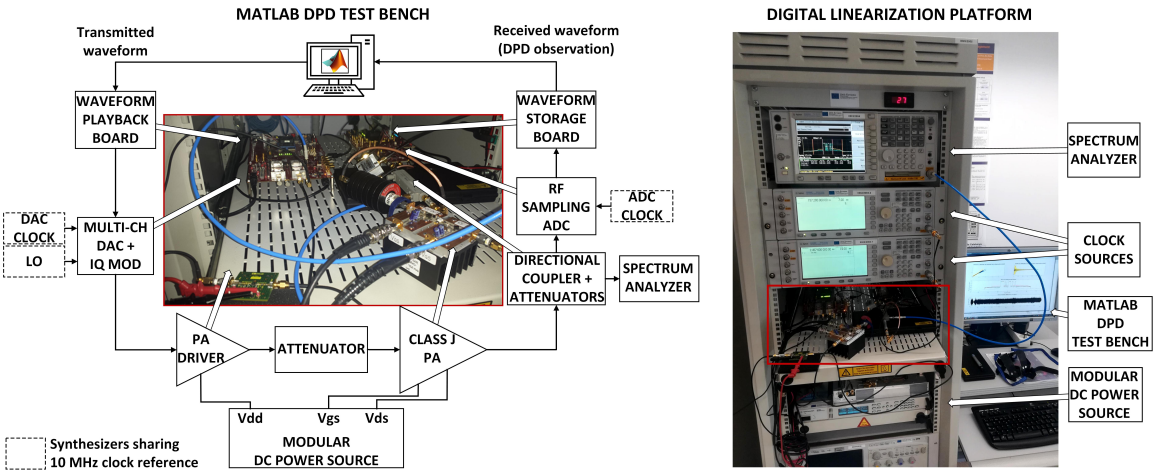


Figure 1.19: SHAPER digital linearization platform for a given test setup.

SHAPER can be used to:

- Extract and validate PA behavioral models or characterize the performance of PA designs.
- Design, simulate, validate and experimentally benchmark polynomial and neural network DPD & CFR algorithms for multiple PA technologies and wireless standards with multi-antenna || channel || band || rate signal configurations.
- Emulate HW constraints (i.e., A/D and D/A bits, data precision) before implementation

and experiment with techniques to reduce HW complexity and computational resources.

- Validate product-oriented FPGA prototyping through MATLAB HW/software cosimulation

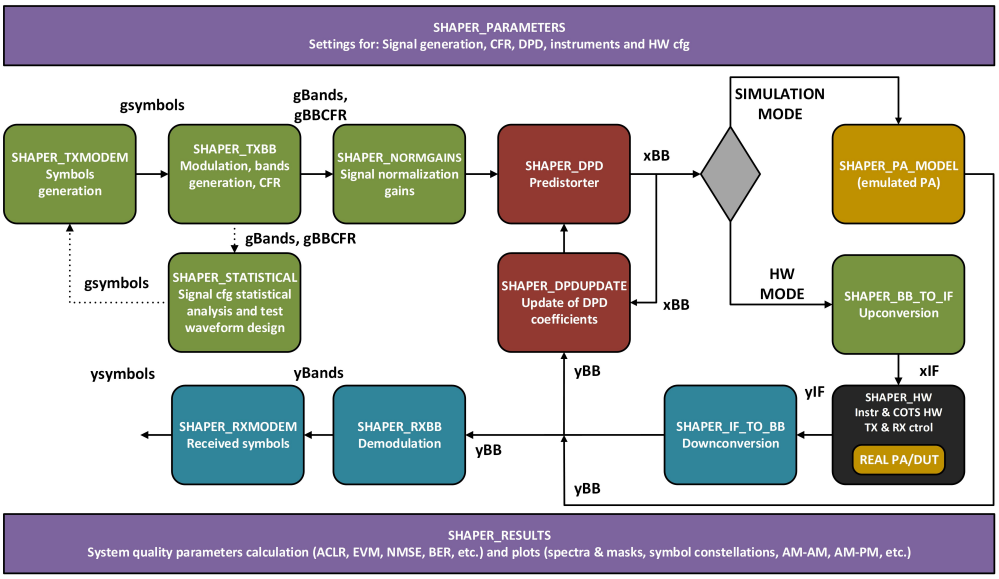


Figure 1.20: SHAPER demonstration platform functional block diagram.

SHAPER features:

- Multi-channel waveform synthesis and acquisition for different BB/BBIQ/IF/RF/mmW DUT interfacing modes reaching up to 1 GHz DPD bandwidth for RF-to-mmWave transceiver operation.
- Key parameter multi-dimensional analysis: i.e. NMSE, ACPR, EVM, raw BER versus output power, crosstalk levels, DPD coefficients, PAPR reduction, complexity reduction, CFR/DPD variants, or any combination of the previous.
- Design of statistically representative experiments and waveforms (Full PAPR statistical characterization of modulated waveform and test waveforms design).
- Remote operation from anywhere in the World, by having a PC with MATLAB and internet connection, thanks to a waveform upload/download file transfer protocol (FTP) MATLAB server/client application.
- The modular MATLAB test bench (shown in Fig. 1.20) has been evolving since the thesis started following a flexible and scalable approach in order to ease the integration of new digital modems and linearization algorithms, and provide different complexity usage levels. It is also featuring full setup and results traceability and high measurement repeatability.

(M5) Since its conception and in parallel to the thesis period, the SHAPER test bed has been used in 4 industrial projects, 4 competitive projects, about 20 publications, 2 PhD thesis, one MSc and two BSc theses.

1.3 Outline

Chapter 1 introduces the mobile communication networks context and the application scenarios considered in this thesis that demand digital linearization (radio access and wireless backhauling). The thesis objectives are then described together with the contributions and novelty linked to these objectives, and other merits of interest.

Chapter 2 provides an insight on the principles behind the digital linearization of power amplifiers. Starting with an overview of the PA operation classes, efficient architectures and classical linearization schemes and linearization metrics, it then follows with describing the adaptive closed-loop DPD and several implementation aspects both for the DPD forward and feedback paths. The chapter ends by listing the most common CFR techniques and showcasing the combined application of CFR and DPD into a spectrum-shared PMR access base station transmitter.

Chapter 3 introduces the main machine learning techniques that are used for robust dimensionality reduction in polynomial SISO digital predistorters (thus reducing the number of DPD coefficients), to prevent ill-conditioning, and to reduce the complexity and number of operations at the DPD identification subsystem. These techniques are also linked to the thesis early prospection and research works. First, histogram-based sample selection techniques that reduce the number of samples to be processed by the DPD system, together with piecewise and long term modeling techniques, are introduced and showcased to guarantee model identification and DPD robustness when targeting a GaN PA operating under 4G-5G burst-like signals with strong memory effects. Second, feature selection techniques like the orthogonal matching pursuit (OMP), that is applied for basis reduction in the forward path, are described. The SW-OMP algorithm that allows trading-off between EVM and ACPR in DPD is also introduced through another use case example. Third, the theory behind feature extraction techniques like the principal components analysis (PCA), that is applied for basis reduction in the feedback path, is also deployed together with some notions on the combination of PCA and partial least squares (PLS). A wireless backhauling use case combining CFR and DPD follows. The thesis builds upon contributing to new and more beneficial variants of the previous techniques for which a relevant contribution to the DPD feedback path is found in the last part of the chapter: A novel independent DPD coefficient estimation technique supported by an adaptive PCA algorithm variant is formulated for SISO polynomial predistorters. The design principles are fully detailed and experimentally validated.

Chapter 4 leverages on the neural network approaches that may be applied to the field of DPD in contrast to the classical polynomial approach. Several general aspects on ANN topologies, nomenclature and formulation are addressed together with more focused design considerations for DPD correction. Specifics on the SISO ANN DPD training and parameter extraction procedures and recommended validation methodologies are also described and demonstrated in a PA ANN behavioral modeling 4G-5G radio access use case involving measured input-output data. Under changing conditions, the adaptive SISO ANN DPD may have strict training time requirements that cannot be met. To address such threat while preserving the transmitter requirements, the chapter ends with a new contribution towards reducing size of the ANN SISO predistorter input dataset. Such technique is based on the combination of new waveform batch selection mechanisms that support higher reduction of DPD training signals' length and so the dataset length (i.e., when compared to multidimensional histogram-based methods) together with an already existing variant of the OMP algorithm that for the first time is applied to select and reduce the number of the ANN input dataset features.

Chapter 5 goes a step forward by targeting both polynomial and neural network digital MIMO predistorters that not only compensate for every transmitter branch RF impairments, but the PA input and output cross talk. After review of the existing state-of-the-art, the MIMO predistorter architectures are fully detailed and depicted. The procedures to combine length reduction mechanisms and machine learning dimensionality reduction techniques are deployed for the three MIMO DPD architectures (polynomial, MISO ANN and MIMO ANN) for any potential combination of the techniques being reported. Three newly applied techniques are considered on top of those described in Chapter 4. The first is UES that removes equations that do not provide useful information for identification and thus reduces the data length. After applying UES, for dimensionality reduction OLS and PCA are proposed for polynomial and ANN MIMO predistorters, respectively. To finish this chapter, an exhaustive benchmark of the proposed techniques demonstrates the gains in terms of performance, complexity and training time.

Finally, Chapter 6 concludes not only discussing the main findings, contribution and benefits of the research conducted, but stating potential limitations and tracing out the direction for future lines of research, according to the evolution of beyond 5G communication networks.

Chapter 2

Power Amplifier Digital linearization Principles

2.1 Power Amplifier Basics and Linearization Overview

2.1.1 RF Power Amplifier Classes

The general input-output operating principles of the PA and the linearity versus efficiency trade-off have been briefly introduced in Section 1.1.2. By choosing the bias point and load condition of an RF PA, such trade-off can be tailored and be optimized either for higher gain, minimum distortion and maximum output power, or to maximize efficiency. In general, the long-standing alphabetical classification of PAs in classes represents the conduction angle of a full cycle of a sinusoidal input signal where the transistor of the amplifier is active and drives current (or conducts) [Wal12].

The most common amplifier classes controlled by the conduction angle are class A, B, AB and C. In these classes the transistors behave as a transconductor. There is a second group of operating classes that are the so-called switching amplifier classes (i.e., with the transistor operating as a switch) where we find class D, E, F, G, S and T. These amplifiers employ digital circuits and pulse width modulation (PWM) to switch between full ON and OFF states driving the output signal harder into saturation to reach higher power efficiency. Assuming ideal input waveform sinusoidal shape and transistors (and ideal biasing circuits and dc blocks), Figure 2.1 shows an ideal operation model for the transistor (upper right corner), how this transistor conducts for a given sinusoidal input and bias point (left side) and which is the amplified version of the input signal according to the loading conditions (right side). It is noted that Vgs stands for gate-to-source voltage, Vds for drain-to-source voltage, Ids is the drain-to-source current and RFin and RFout are the input modulated signal and the output amplified version, respectively.

The power amplifier classes are summarized in the following:

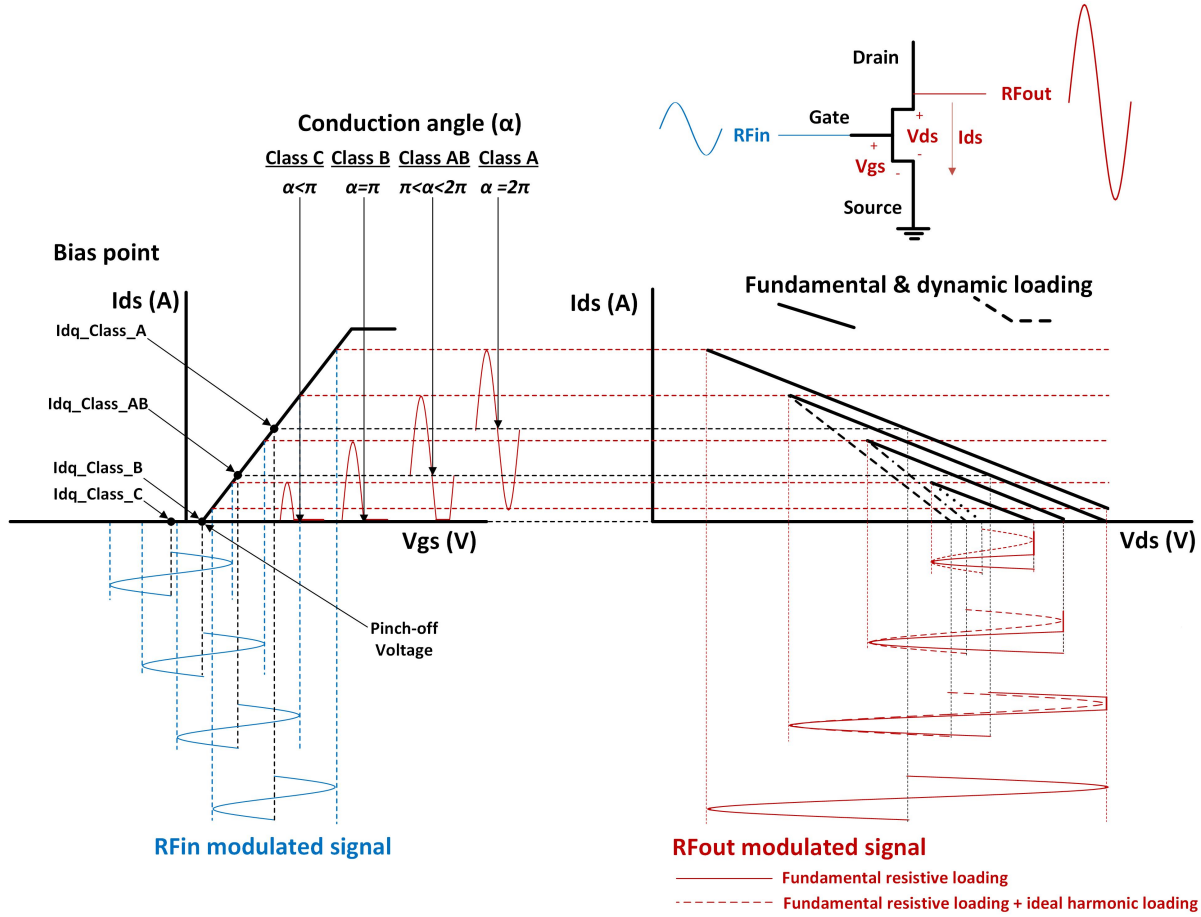


Figure 2.1: RF power amplification mechanism for operation classes A, AB, B and C.

Class-A Amplifiers: As depicted in Figure 2.1-left, these power amplifiers operate in full angle conduction (2π or 360°) and thus drive current and power 100% of the time. The input signal drive level is also kept low to avoid entering the pinch-off region. These are the highest linearity and gain power amplifiers but their efficiency is low and reaches a maximum theoretical value, obtained when operating with continuous-wave (CW) signals, of 50%. As previously introduced, the higher the linearity the lower the efficiency is (and vice versa) and the higher the amount of power dissipated in the form of heat. These amplifiers therefore need heat sinks which add size, weight, volume and cost.

Class-B Amplifiers: These amplifiers were proposed to cope with the low-efficiency and heating drawbacks of class-A amplifiers. Class B only operates at half of the full conduction angle or 180° by biasing the transistor gate at the pinch-off point where the transistor turns off and the quiescent drain current is zero. The efficiency of these power amplifiers, that conduct half of the time than class A, is higher reaching a theoretical maximum of 78.5%. The linearity of these amplifiers, however, is worse than that of class A when using a single transistor. The harmonic distortion that will be presented by the PA, which produces half-wave rectified output sinusoidal

waveforms, needs to be filtered out. Typically, an RF output network presenting resistive loading at the fundamental component and short circuiting the even-harmonic components is employed at the output of the PA. Class-B amplifiers require inputs signals twice as larger as class-A amplifiers to maximize the usage of the linear region of the transistor (have lower gain). Another typical configuration is the "push-pull" amplifier that combines two transistors that are used to supply current alternately during the positive and negative input signal half-cycles and combines the two waveforms being generated at the output of the amplifier to reconstruct an amplified version of the original input signal.

Class-AB Amplifiers: These amplifiers combine characteristics from both class-A and class-B operation to gain the benefits from both. The transistor is biased to a quiescent point between the pinch-off point and the class-A bias point. The conduction angle is now higher than 180° but less than 360° to minimize the class-B crossover distortion produced when shooting off at 180°. The efficiency can be traded off and be adjusted where needed between that of class B and class A. The clipping of the RF voltage will be also producing odd-harmonic distortion for which the use of harmonic loading may also be required to enhance linearity.

Class-C Amplifiers: These amplifiers are in the border between the transconductor and the switching-mode amplifier groups. Class-C results from biasing the transistor below pinch off and therefore it will be active or driving current for less than half of the cycle (i.e., below 180°), conducting only for short pulse-like periods and requiring high input drive levels, and limiting its usability to specialized applications. While class A, AB and B are considered linear amplifiers since the output signals amplitude and phase is linearly related to the input signal ones, class C is more of an improved efficiency class targeting beyond 80% efficiency with a maximum theoretical level of 100%. However, the quest for the highest efficiency comes at the expense of a much lower gain and nonlinear behavior for which typical traded-off efficiencies are around 85%. In a class-C amplifier the harmonic levels escalate rapidly when going below 180° conduction, both for odd an even harmonics, which poses a much greater challenge to fulfill the necessary harmonic loading conditions to have a sinusoidal output signal. Despite the drawbacks and limited applicability, it is worth mentioning that class-C amplifiers have found a relevant niche in the RF amplifiers employed in base stations by acting as the peaking amplifier in Doherty PA architectures.

The switching-mode amplifier classes (i.e., from class C to S) are hardly suitable for linear application like that required by radio access or wireless backhaul infrastructure due to introducing significant out-of-band distortion thus impairing the spectrum adjacent channels. A comprehensive description of such operating modes can be found in [Wal12]. These amplifiers, however, can be employed as building blocks of other efficient PA architectures like those that are described in Section 2.1.2, or be used in conjunction with conveniently adapted versions of the linearization schemes deployed in Section 2.1.3. While some class-AB amplifiers have been

digitally linearized in the early research works, for the main contributions of the thesis a more power efficient alternative based on the class-J power amplifier described in [Riz11] which can be seen as a class AB using capacitive harmonic termination, has been combined for validation of the ML-assisted DPD techniques being proposed. To describe the main operating principles of class-J amplifiers it is convenient to introduce first how does class F operate. The two classes are described as follows:

Class-F Amplifiers: Class F provides the means to boosting the maximum efficiency of a class-B or deep class-AB amplifier by more than 10% by using properly tuned harmonic resonators in the output network to shape the drain waveforms. This is first accomplished by adding a third harmonic component (ideally, all higher-order odd harmonics) to the input voltage signal, i.e., driving the transistors to saturation, which will contribute to flatten the output voltage signal and convert it into a square-wave output signal. The presence of a third-harmonic component allows in addition to have higher amplitude fundamental component at the output signal. Second, the output matching or loading network will now i) transform the fundamental load line resistance to the termination impedance (i.e., being typically 50 ohms), ii) short-circuit the second harmonic (and ideally higher order even harmonics), iii) present typically a high impedance to the odd harmonics. Despite Class-F amplifiers may reach 100% efficiency in theory, about 88% when focusing only up to the third harmonic, in practice it will be hard to have open circuit termination for higher frequencies and higher power devices because of the escalation in the output capacitance of the active device.

Class-J Amplifiers: A class-J amplifier may be seen as a variant of a class-F amplifier but now using the second harmonic (instead of the third one) to shape the output waveforms. This operating mode was promoted by Cripps in [Cri06] following the works in [Raa01] and [Rho03].The output class J voltage waveform requires i) a fundamental load consisting of the regular load-line resistive component together with an equal reactive component to keep the voltage wave above zero (complex loading is required at the fundamental to conveniently shape the output wave voltage), ii) a second harmonic reactance load with comparable value to the load-line resistance. While the efficiency may be similar to that of the class B amplifier, class J has the advantage (over either class B or class AB counterparts) that the second harmonic is not demanding a short-circuit termination but a capacitive reactance that, in some cases, may be provided by the output capacitance of the transistor device. Class-F amplifiers may present significant problems when the performance needs to be kept over a considerable bandwidth due to having to short-circuit the even harmonics. This is not the case of the class-J amplifier which may provide, over a broad frequency range, both enhanced efficiency and similar linearity as class-AB or class B amplifiers when operating with high-PAPR input signals (i.e., due to the absence of resonant impedance conditions).

2.1.2 Efficient Power Amplifier Architectures

As outlined in the previous Chapter, the PA is not only one of the main power-hungry elements in the transmitter but also one of the main sources responsible for introducing nonlinear distortion. The main drawback of modern communication signals (mainly OFDM-based) is that the cost of a better use of the spectral resources is an increase of the PAPR of the signals. Again, this constraint forces the system to operate the PA at a large output power back-off to guarantee the required linearity level at the antenna port, leading to a serious degradation of average power efficiency. To minimize power dissipation, highly efficient amplification architectures based on the dynamic load or dynamic supply modulation have been proposed in the literature. Some of the most popular solutions are envelope tracking PAs [Pop17], Doherty PAs [Pen16b], load-modulated balanced amplifiers (LMBAs) [Qua22], and linear amplification with nonlinear components (LINC) or outphasing PAs [Bar16]. Typically, these amplification topologies are designed targeting high average efficiency figures over bandwidth and power back-off, however, the linearity is left as a problem to be addressed at system level by using some kind of linearizer.

Therefore, these topologies require the use of digital signal processing to generate additional signals, including the supply or load modulated control signal, to maximize power efficiency, as well as digital predistortion (DPD) linearization techniques to guarantee the stringent linearity requirements of today's systems, especially with the increasing signal bandwidth.

Doherty Power Amplifier

For base stations, for example, one popular solution is the use of Doherty PAs, proposed by W. H. Doherty in early 1936 [Doh36]. The typical Doherty architecture consists of two PAs: a class-AB carrier PA and a class-C peaking PA. The operating principle of Doherty PA is as follows. At low input power level, the peaking PA is turned off, only the carrier PA amplifies the input signal. When the input power level increases up to a threshold (i.e. the point where the carrier PA starts to go into compression) the peaking PA is turned on to amplify the input signal together with the carrier PA. The Doherty PA suffers from nonlinear distortion, mainly due the gain compression that appear just before the peaking amplifier starts conducting. This phenomenon of double gain compression can be compensated using DPD linearization [Kim06].

Load-Modulated Balanced Amplifier

The LMBA, proposed by Shepphard et al. in [She16], is based on a balanced PA where a control signal power is injected at the isolated port of the output 90° coupler modulating the load at each balanced device. Unlike in the Doherty PA, in the LMBA the control signal power recovery is independent to the load modulation and it can be fully recovered at the output of

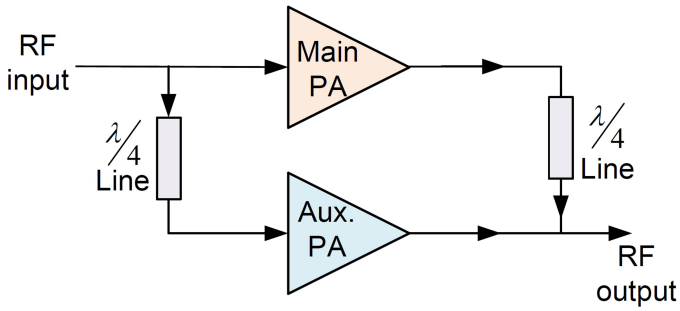


Figure 2.2: Block diagram of the Doherty PA.

the LMBA. Another key property of the LMBA is its wider bandwidth of operation, since the load modulation can be applied for the whole frequency band of the 90° coupler, that is normally larger than the bandwidth of a Doherty combiner. The additional degrees of freedom offered by a dual-input LMBA can be used to optimize the performance on the same or larger bandwidth, or to improve other performance metrics such as average power efficiency and linearity [Wan21b, Gui22], when properly combined with DPD linearization.

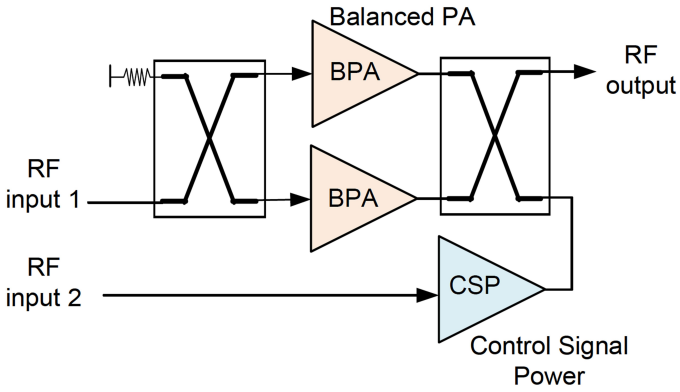


Figure 2.3: Block diagram of the LMBA PA.

Outphasing or LINC Power Amplifier

The outphasing PA, introduced by H. Chireix in [Chi35] circa 1935, consisted in using phase control of two constituent branch PAs operated in saturation by enabling constant-output envelope signals that were summed at the output of the system to allow amplitude modulation. Later in 1974, D.C. Cox [Cox74] generalized the LINC approach. The main difference is that, unlike the non-isolating Chireix combiner (providing good efficiency but not-so-good linearity), the LINC combiner isolates the two PA outputs, which improves the combining linearity at the price of high-power dissipation at high outphasing angles. The performance of an outphasing architecture is highly sensitive to frequency variations and thus there is a trade-off between the achievable bandwidth and highly efficient power control range. To deal with the nonlin-

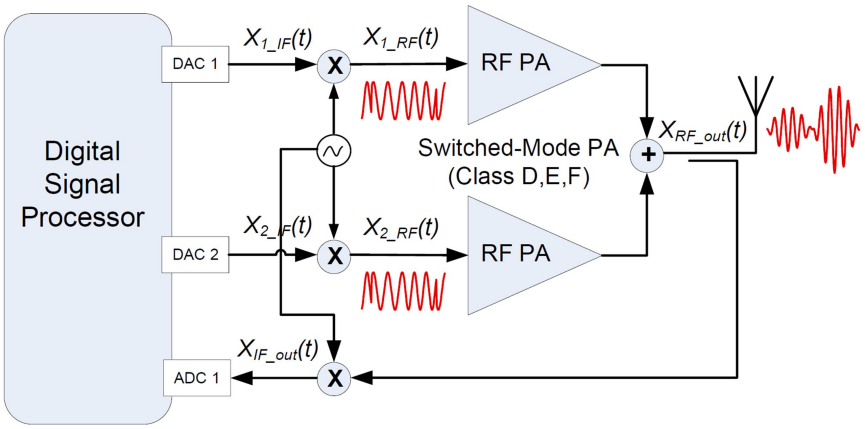


Figure 2.4: Block diagram of the outphasing or LINC PA.

ear distortion introduced by the non-isolating Chireix combiner, hybrid modes of operation are combined with DPD to guarantee the required linearity levels while trying to keep the average power efficiency as high as possible [Gil20c].

Envelope Tracking Power Amplifier

In PAs with dynamic supply modulation, the supply voltage of the RF PA is adjusted according to the envelope of the RF carrier. One of the precursors of dynamic supply modulation is the envelope elimination and restoration (EE&R) technique introduced by L. R. Kahn in [Kah52] circa 1952. Unlike current ET PAs, where the RF signal is amplitude and phase modulated, the EE&R technique operates a high-efficiency switched-mode RF PA with a constant-amplitude phase-modulated RF signal (the amplitude information of the RF signal at the output of the EE&R is added through the dynamic supply of the RF PA). ET PAs have been commercially exploited mainly for handset applications. Thanks to the dynamic supply, the RF PA (linear current-source mode, Doherty, or even LINC PAs) can be forced to operate, all the time and for any non constant envelope signal, close to saturation which increases the power efficiency for high PAPR signals. Several strategies can be designed to shape the supply voltage signal to achieve better linearity and efficiency or to meet the slew-rate and bandwidth restrictions of the envelope tracker modulator. When considering the amplification of wide bandwidth signals, the use of linearization approaches in the envelope supply path (to compensate for the non-ideal behavior of the envelope tracker modulator) or in the I-Q signal path (e.g., DPD to compensate for the distortion introduced by RF PA) results necessary [Gil12]. A combined linearization strategy is proposed in [Li23] to compensate for the envelope tracker modulator leakage in the supply path and the PA nonlinear distortion in the RF path, when considering signals up to 200 MHz.

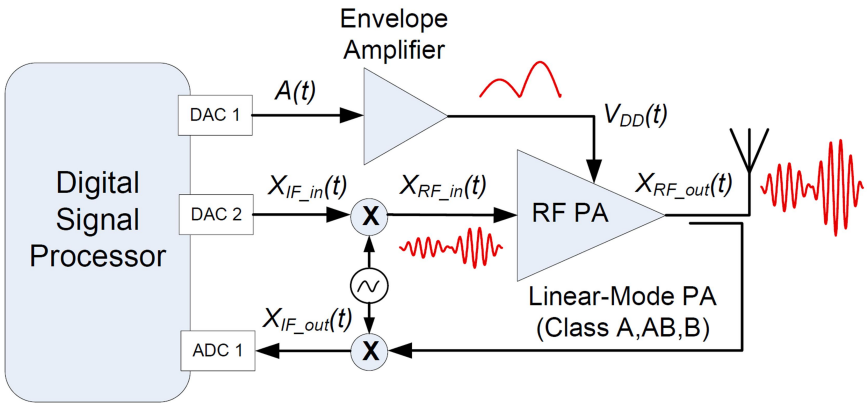


Figure 2.5: Block diagram of the ET PA.

2.1.3 Classical Linearization Schemes

Feedforward Linearization

The feedforward linearizer, that was proposed by Harold S. Black in 1923 at Bell Telephone Laboratories, is considered to be the first linearization topology and, until the irruption of digital predistortion, it was used in communications applications for operating with wideband signals. Fig. 2.6 shows the block diagram of a general Feedforward linearizer.

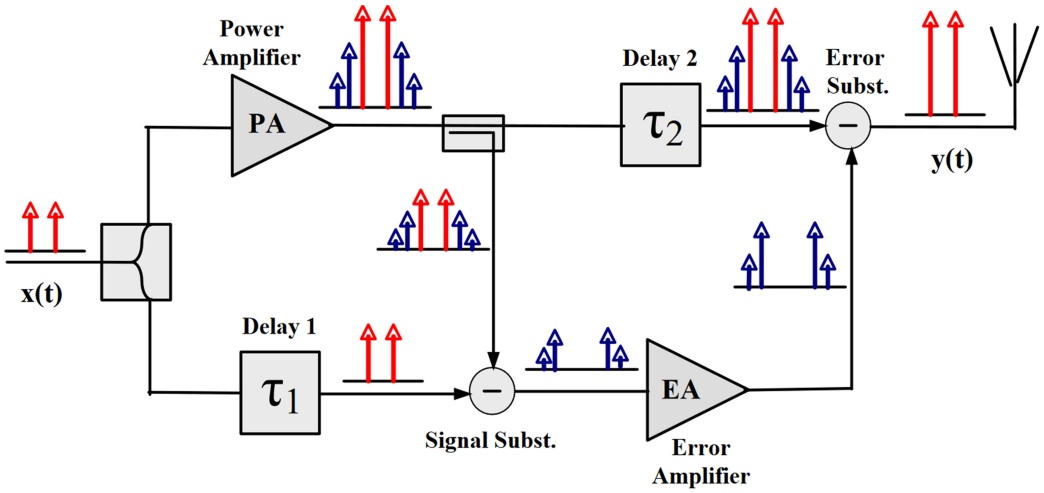


Figure 2.6: Simplified feedforward block diagram and principles of operation [Gil07a].

The feedforward functioning is quite straight-forward. Considering a two-tone test, as depicted in Fig. 2.6, the input signal is first equally split and fed to the upper and lower paths, respectively. The signal in the upper path is amplified by the PA. The output signal of the PA shows intermodulation distortion products. An attenuated sample of the distorted PA output signal is fed to the lower path and then subtracted with a delayed version of the input signal. Therefore, considering an ideal match between the lower path delay (τ_1) and the delay introduced

by the PA (in the upper path), the resulting error signal consists only of the intermodulation distortion products caused by the PA. Then, the error signal is linearly amplified by the error amplifier and fed to the upper path. Contemporary, the PA output signal in the upper path is delayed (τ_2) by an amount equal to the delay introduced by the error amplifier. Finally, if the cancellation has been perfectly done at the output of the error subtracter, the amplified version of the desired original signal will be delivered.

One of the advantages of the feedforward linearizer is that, in contrast to the later feedback linearizer, it is unconditionally stable and can be used for wideband signals. However, its open-loop nature makes it too sensitive to delay mismatches and tolerances of the components, which ultimately may degrade the linearization performance [Pot99]. Black tried to combat these effects but, after four years working on it, he could not maintain the required cancellation over temperature, aging, and frequency with the vacuum tube amplifier technology available at that time [Kat16] Over time, the concept was revived with the arrival of the solid-state power amplifiers and the need for highly linear PAs for cellular telephone base stations in the 1980s. With adaptive control, some reasonable cancellation could be guaranteed under temperature variation and aging but, in practice, such cancellation could be only achieved up to about 6 dB from saturation to allow the auxiliary amplifier in the lower path to operate in a linear regime (so as not to introduce distortion on its own). The feedforward linearizer can take to significant power efficiency degradation (even when linearity levels are maintained [Gil04]). The complexity overhead and increased power consumption are relevant disadvantages that have been limiting the deployment of feedforward linearization solutions.

Feedback Linearization

The rationale behind feedback linearization consists in feeding the output back to the input, 180° out of phase, reducing the distortion at the expense of gain. Harold S. Black demonstrated the benefits of negative feedback [Bla34] already in 1927. Since then, feedback theory has been developed and applied in many applications from control to communications. A general feedback closed-loop block diagram including input and output additive disturbances is shown in Fig. 2.7.

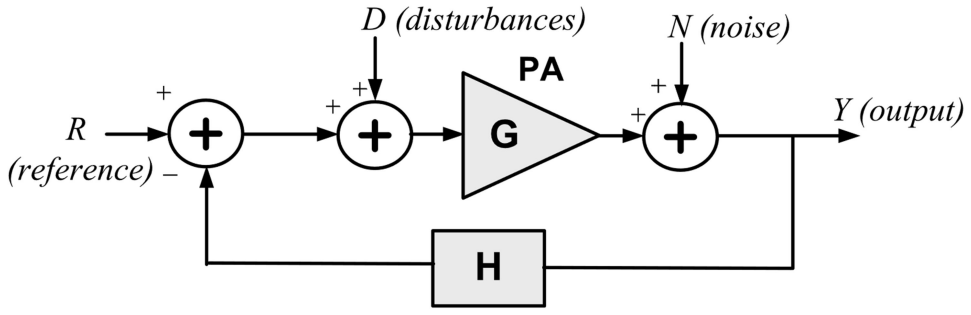


Figure 2.7: General block diagram of feedback linearization technique [Gil07a].

Considering the generic notation in Fig. 2.7, the output of the feedback system can be expressed as

$$Y(\omega) = \frac{G(\omega)}{1 + G(\omega)H(\omega)} R(\omega) + \frac{G(\omega)}{1 + G(\omega)H(\omega)} D(\omega) + \frac{1}{1 + G(\omega)H(\omega)} N(\omega) \tag{2.1}$$

Where $R(\omega)$ is the reference signal (input signal), $Y(\omega)$ the output signal, $D(\omega)$ the input disturbance and $N(\omega)$ the output disturbance (noise), all of them defined in the frequency domain. As described in (2.1), the bigger is $H(\omega)$, the more insensitive is the feedback system to input disturbances. Moreover, the bigger is $G(\omega)$ or $H(\omega)$, the more insensitive the output noise is. However, the price to reduce input disturbances is a loss in the overall system gain (amplification).

Some feedback structures have been proposed for communications systems, such as Cartesian feedback linearizers [Boo11] , polar loop feedback [Sow04] or envelope feedback [Che02] linearizers. The approach may provide relevant intermodulation distortion (IMD) rejection but is limited in terms of high-frequency operation. Other feedback linearizer drawbacks, beyond the loss of amplification gain and high-frequency operation, is that the linearization performance is degraded when considering wideband signals (i.e., there is a practical difficulty in making a feedback system responding to signal envelopes faster than several tens of MHz due to the required delay for the PA and rest of the components), and that may suffer from stability issues.

Predistortion Linearization

The quest for simplicity and cost effectiveness, wideband operation and dynamic range and has promoted the use of predistortion linearizers which neither have loop nor delay issues and are operated before the PA meaning the process is less energy hungry. According to [Kat16], the predistortion linearization technique was introduced by Adel A. M. Saleh and J. Salz from Nokia Bell Labs in 1983 [Sal83]. The principle of predistortion, independently if it is analogue or digital, is quite straightforward. As depicted in Fig. 2.8, it consists in preceding the PA with a device called predistorter in order to counteract the nonlinear characteristic of the PA. The objective of the predistorter is then to ideally reproduce the inverse PA nonlinear behavior and as a result having linear amplification at the output of the PA.

Analog predistortion (APD) uses an analog circuit to predistort the input signal. One of the advantages of ADP is that the bandwidth expansion is taking place in the analog domain, which relaxes the clock rate requirements in the digital-to-analog converter (DAC) to generate the analog transmitted signal. For the design of analog predistorters, the key point is to tune a nonlinear circuit that has to show the exact inverse nonlinear characteristic as the targeted PA. Several solutions have been proposed to generate the inverse characteristic using analog circuitry [Gao18, Gum18].

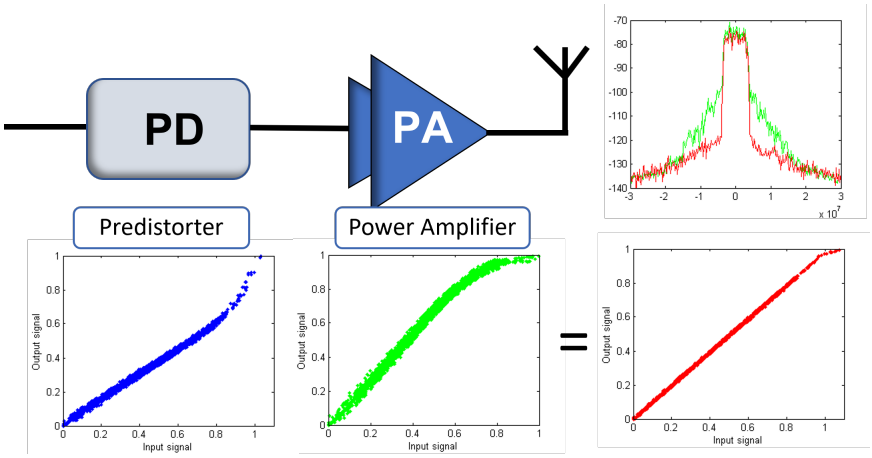


Figure 2.8: The concept of predistortion.

Since the 1980s, with the development of signal processing hardware (e.g. system-on-chip (SoC) and field programmable gate arrays (FPGAs) with high-speed processing and memory capabilities, digital predistortion (DPD) linearization has become the dominant linearization technique [Kat16]. In DPD the predistortion of the transmitted signal is carried out at base-band in a digital signal processor. By applying DPD to the complex digital base-band signal, the predistorter signal to be transmitted suffers a bandwidth expansion (e.g., 3× to 5× the original signal bandwidth) that has to be taken into account to determine the DAC clock rate. In addition, a feedback or observation path is necessary to monitor the amplified signal and estimate the coefficients of the DPD.

DPD linearizers are nowadays a fundamental building block vastly adopted by today's base station modems and user handsets since all its signal processing can be achieved in a DSP simplifying the RF layout. Analog predistortion (APD) linearizers are more cost-effective but since their performance is lower than DPD, their use is more limited to applications requiring large bandwidth where DPD is technically unfeasible or too costly. The increasing PAPR and the quest for efficiency, has made the combination of Doherty or ET/polar PAs and DPD a *de facto* standard.

The increasing signal bandwidth evolution at every mobile communications technology generation (i.e. from a few kHz or hundreds of kHz to tens and hundreds of MHz) has also given increasing importance to the mitigation of the memory effects being produced in the PA. Such effects make the amplitude of the output signal not only be dependent on the actual input signal amplitude but also on several past values. These memory effects are caused by different physical mechanisms like the change of nonlinearities of the PA and the behavior of the impedance of decoupling networks with frequency (the latter with respect to the envelope frequency), changes in the active devices biasing due to the envelope variation, or the impact of electrothermal effects in the device operation. Further detail on the basic principles of DPD linearization compensating

for the memory effects is given in Section 2.2.

Linearization performance metrics

The most common metrics used to evaluate the linearization performance are listed in the following. To quantify the out-of-band distortion, the adjacent channel power ratio (ACPR) is commonly used. The ACPR is in this thesis defined as the ratio of the power delivered in the adjacent channels (adj) to the total power over the channel bandwidth (ch).

$$\text{ACPR (dB)} = 10\ \log_{10}\left(\frac{\int\limits_{\text{adj}} P_{out}\,(f)\ df}{\int\limits_{\text{ch}} P_{out}\,(f)\ df}\right) = 10\ \log_{10}\left(\frac{\int\limits_{\text{adj}} |\ Y(f)\ |^2\ df}{\int\limits_{\text{ch}} |\ Y(f)\ |^2\ df}\right) \tag{2.2}$$

where $Y(f)$ is the Fourier transform of the measured PA output signal. Consequently, taking into account the logarithmic scale, the more negative the ACPR is, the less spectral regrowth and the more linear is the PA. The ACPR is calculated for both the upper sideband (US) and the lower sideband (LS), and a common approach is to report the worst case ACPR, i.e., $\text{ACPR}_{\text{worst}} = \max\{\text{ACPR}_{\text{LS}}, \text{ACPR}_{\text{US}}\}$.

The in-band distortion introduced by the PA can be observed in digital linear modulations such as the quadrature amplitude modulation (QAM) schemes. It can be quantified with the error vector magnitude (EVM) figure of merit. The EVM measures the effects of the distortion on the amplitude and phase (or the In-phase and Quadrature components) of the modulated output and it is defined as the square root of the ratio of the mean error vector power to the reference (ref) power expressed as a percentage (%),

$$\text{EVM (\%)} = \sqrt{\frac{\sum\limits_{n=1}^{L}\left(\Delta I^2[n] + \Delta Q^2[n]\right)}{\sum\limits_{n=1}^{L}\left(I^2_{\text{ref}}[n] + Q^2_{\text{ref}}[n]\right)}} \cdot 100 \tag{2.3}$$

where $\Delta I[n] = I_{\text{ref}}[n] - I_{\text{meas}}[n]$, $\Delta Q[n] = Q_{\text{ref}}[n] - Q_{\text{meas}}[n]$, with $I_{\text{meas}}[n]$ and $Q_{\text{meas}}[n]$ being the in-phase ($I_{\text{meas}}[n] = Re\{y_{\text{demod}}[n]\}$) and quadrature ($Q_{\text{meas}}[n] = Im\{y_{\text{demod}}[n]\}$) components of the measured demodulated PA output signal ($y_{\text{demod}}[n]$), respectively. Finally, L is the number of samples. The EVM includes information on the transmit filter accuracy, D/A-converter, modulator imbalances, untracked phase noise, and power amplifier non-linearity. In a similar manner to the spectral regrowth limitations, communications standards determine maximum levels of the EVM permitted at the transmitter antenna and at the receiver, depending on the modulation scheme used and the use (or not) of any codification.

One of the most commonly used PA or DPD modeling performance indicators is the normalized mean squared error (NMSE). The NMSE, for example, can be used to indicate how well the model is approximating the reality, i.e., it measures the resemblance of the discrete-time, complex-valued estimated output ($\widehat{y}[n]$) and the discrete-time, complex-valued measured output of the PA ($y[n]$). It is therefore driven by the in-band channel error which has considerably more power than that between out-of-band adjacent and alternate channel error).

$$\text{NMSE (dB)} = 10\ \log_{10}\left(\frac{\sum\limits_{n=1}^{L}|y[n]-\widehat{y}[n]|^2}{\sum\limits_{n=1}^{L}|y[n]|^2}\right) \tag{2.4}$$

Being inspired by the ACPR metric, and in contrast to the NMSE metric, the adjacent channel error power ratio (ACEPR) is oriented at highlighting the modeling capabilities of the nonlinear part (impacting in the out-of-band error power), most of the times masked by the dominant linear part. Thus, the ACEPR calculates the error signal power in the adjacent and/or alternate channels relative to the in-band channel power of the measured signal as

$$\text{ACEPR}\ (dB) = 10\ \log_{10}\left(\frac{\int\limits_{\text{adj}}|E(f)|^2\ df}{\int\limits_{\text{ch}}|Y(f)|^2\ df}\right) = 10\log_{10}\left(\frac{\int\limits_{\text{adj}}|\ Y(f)-\widehat{Y}(f)\ |^2\ df}{\int\limits_{\text{ch}}|\ Y(f)\ |^2\ df}\right) \tag{2.5}$$

where $E(f)$ is the Fourier Transform of $e(n) = y(n) - \widehat{y}(n)$, being $y(n)$ and $\widehat{y}(n)$ the measured output signal and its modeled or estimated signal, and where $Y(f)$ and $\widehat{Y}(f)$ are the Fourier Transforms of such signals.

2.2 Digital Predistortion Linearization

2.2.1 Power Amplifier Behavioral Modeling

Power amplifier behavioral models are mathematical descriptors of the nonlinear behavior and memory effects of PAs. Unlike physical models, where it is necessary to know the electronic elements that comprise the PA, their constitutive relations and the theoretical rules describing their interactions, the extraction of PA behavioral models relies only on a set of input-output observations. Consequently, their accuracy is highly sensitive to the adopted model structure and the parameter extraction procedure. In general, the same basis functions used for approximating the response of the PA are also used to estimate its inverse response.

Several PA behavioral models can be found in the literature addressing not only the nonlinear modeling of SISO systems but also MISO systems. For example, when having to characterize concurrent multi-band transmissions or dynamic supply modulation strategies for the PA [Gil19].

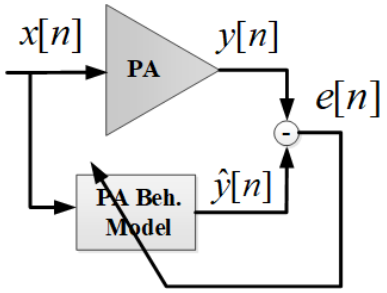


Figure 2.9: Identification of the power amplifier behavior.

Some of the most commonly used polynomial-based behavioral models can be seen as a simplified approximations of the general Volterra series. Volterra series are aimed at describing time-invariant nonlinear systems with fading memory. Following the notation in Fig. 2.9, the discrete-time low-pass equivalent Volterra series formulation considering complex-valued signals is described as follows,

$$\hat{y}[n] = \sum_{p=1}^{P} \sum_{q_1=0}^{Q-1} \sum_{q_2=q_1}^{Q-1} \cdots \sum_{q_p=q_{p-1}}^{Q-1} \cdots \\ \cdots \sum_{q_{2p-1}=q_{2p-2}}^{Q-1} h_{2p-1}(q_1, q_2 \cdots, q_{2p-1}) \prod_{i=1}^{p} x[n-q_i] \prod_{j=p+1}^{2p-1} x^*[n-q_j] \tag{2.6}$$

where $x[n] \in \mathbb{C}$ and $\hat{y}[n] \in \mathbb{C}$ are the PA input and estimated PA output signals, respectively. The series is composed by P kernels of increasing dimensional order. The main drawback of using the full Volterra series is that the number of parameters grows exponentially when considering higher order kernels and typical communication signals do not present enough richness to fully excite these kernels, which ultimately may lead to an ill-conditioned problem.

Alternatively, one of the most widely used models in literature is the generalized memory polynomial (GMP), proposed by Morgan et al. in [Mor06]. The estimated PA output using the GMP is defined as follows,

$$\begin{aligned} \hat{y}[n] = & \sum_{i=0}^{N_a-1} \sum_{p=0}^{P_a-1} \alpha_{p,i} \cdot x[n-\tau_i^a] \left| x[n-\tau_i^a] \right|^p + \\ & \sum_{j=1}^{M_b} \sum_{i=0}^{N_b-1} \sum_{p=1}^{P_b} \beta_{p,i,j} \cdot x[n-\tau_i^b] \left| x[n-\tau_i^b-\tau_j^b] \right|^p + \\ & \sum_{j=1}^{M_c} \sum_{i=0}^{N_c-1} \sum_{p=1}^{P_c} \gamma_{p,i,j} \cdot x[n-\tau_i^c] \left| x[n-\tau_i^c+\tau_j^c] \right|^p \end{aligned} \tag{2.7}$$

where α_{pi}, β_{pij} and γ_{pij} are the complex coefficients describing the model, and τ^a, τ^b and τ^c (with $\tau \in \mathbb{Z}$ and $\tau_0 = 0$) are the most significant non-consecutive delays of the input signal

$x[n]$ that better contribute to characterize memory effects. The total number of coefficients is $O = P_aN_a+P_bN_bM_b+P_cN_cM_c$. The GMP model is an evolved version of the memory polynomial (MP) model in [Kim01] (appearing as the first summand in the previous equation and featuring P_aN_a coefficients) that has bi-dimensional kernels considering cross-term products between the complex signal and the lagging and leading envelope terms. This increases the accuracy of the modeling at the price of increasing the number of parameters when compared to MP.

There are plenty of other behavioral models in literature used for DPD purposes in SISO systems, just to mention a couple of examples, the NARMA model proposed in [Mon07] or the Dynamic Deviation Reduction (DDR) Volterra series in [Zhu06]. Further information on PA behavioral models for SISO systems can be found in [Sch09]. In addition, when considering concurrent multi-band transmissions such as in [Rob13], or even combined with PA dynamic supply modulation strategies such as in [Gil15], or also in multi-antenna systems where each transmit path has its own PA and antenna element such as in [Hau18], MISO behavioral models are required to characterize the different sources of nonlinear behavior. In addition, as an alternative to polynomial-based behavioral models, ANN and SVR approaches have been used in literature for PA behavioral modeling and DPD linearization purposes. Some of these behavioral models will be presented and discussed in detail in the following Chapters of this Thesis.

In general, the estimated PA behavioral model output $\hat{y}[n]$ (for $n = 0, 1, \cdots, L-1$), can be defined following a matrix notation as

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{w} \tag{2.8}$$

where $\boldsymbol{w} = \Big(w_1, \cdots, w_i, \cdots, w_N\Big)^T$ is the $N \times 1$ vector of parameters (or PA modeling coefficients) and $\boldsymbol{X}$ is the $L \times N$ data matrix (with $L \gg N$) containing the basis functions or components. The data matrix can be defined as

$$\boldsymbol{X} = \Big(\boldsymbol{\phi_x}[0], \boldsymbol{\phi_x}[1], \cdots, \boldsymbol{\phi_x}[L-1]\Big)^T \tag{2.9}$$

where $\boldsymbol{\phi_x}[n] = \Big(\varphi_1^x[n], \cdots, \varphi_i^x[n], \cdots, \varphi_N^x[n]\Big)^T$ is the $N \times 1$ vector of basis functions $\varphi_i^x[n]$ (with $i = 1, \cdots N$) at time n. Therefore,

$$\hat{y}[n] = \boldsymbol{\phi_x}^T[n]w[n]. \tag{2.10}$$

The general equation in (2.8) can be particularized for any behavioral model or subset of basis functions. In general, the problem has no exact solution since it is over-determined (i.e. it has more equations than unknowns). To identify the vector of coefficients $\boldsymbol{w}$ we define a cost function that takes into account the identification error $\boldsymbol{e}$ expressed, as depicted in Figure 2.9, as

$$\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}. \tag{2.11}$$

Taking the ℓ_2-norm squared of the identification error, the least squares (LS) minimization problem can be defined as

$$\min_{\boldsymbol{w}} \|\boldsymbol{e}\|_2^2 = \min_{\boldsymbol{w}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2. \tag{2.12}$$

Taking the derivative of the cost function $J(\boldsymbol{w}) = \|\boldsymbol{e}\|_2^2$ and setting it to zero, it can be proved that the solution to the least squares (LS) minimization problem in (2.12) is given by

$$\boldsymbol{w} = (\boldsymbol{X}^H\boldsymbol{X})^{-1}\boldsymbol{X}^H\boldsymbol{y} \tag{2.13}$$

The most common numerical methods used to solve the LS problem are detailed in Section 2.2.3.

2.2.2 Closed-Loop Digital Predistortion

The block diagram in Fig 2.10 shows the DPD function and the closed-loop path for coefficients adaptation following a direct learning approach. The input-output relationship at the DPD block in the forward path is

$$x[n] = u[n] - d[n] \tag{2.14}$$

where $x[n]$ is the signal at the output of the DPD block, $u[n]$ is the input signal, and $d[n]$ is the distortion signal which can described using a particular PA behavioral model as per (2.7) for the GMP model. In such a case, the PA behavioral model or basis functions will be reused to model the distortion signal since they will share the same nonlinear nature, now having different coefficients to model distortion instead of overall PA response.

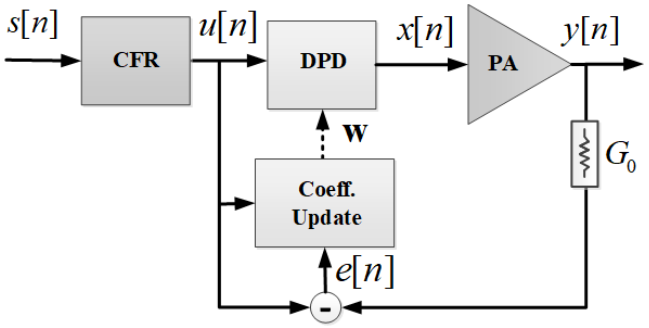


Figure 2.10: Block diagram of CFR and DPD with direct learning DPD adaptation.

Following a matrix notation, we can rewrite (2.14) as

$$\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{U}\boldsymbol{w}, \tag{2.15}$$

where $\boldsymbol{u}$ is the $L \times 1$ input vector and L is the number of samples (i.e., $n = 0, 1, \cdots, L-1$), $\boldsymbol{x}$ is the $L \times 1$ predistorted vector and $\boldsymbol{w}$ now is the $N \times 1$ vector of DPD coefficients. The $L \times N$ data matrix $\boldsymbol{U}$ containing the basis functions is defined as

$$\boldsymbol{U} = \Big(\boldsymbol{\phi}_{\boldsymbol{u}}[0], \boldsymbol{\phi}_{\boldsymbol{u}}[1], \cdots, \boldsymbol{\phi}_{\boldsymbol{u}}[L-1]\Big)^T \tag{2.16}$$

where $\boldsymbol{\phi_u}[n] = \left(\varphi_1^u[n], \cdots, \varphi_i^u[n], \cdots, \varphi_N^u[n]\right)^T$ is the $N \times 1$ vector of basis functions $\varphi_j^u[n]$, for $j = 1, \cdots, N$, at time n. Therefore,

$$x[n] = u[n] - \boldsymbol{\phi_u}^T[n]w[n]. \tag{2.17}$$

Again, these basis functions can be particularized by considering any behavioral model, such as, for example, the GMP model in (2.7).

A direct learning approach [Bra08] is considered (see Fig. 2.10) to identify the coefficients of the DPD model. Applying an iterative learning approach, the DPD coefficients at the k^{th} training iteration can be calculated as follows,

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k + \mu\ \boldsymbol{\Delta w}. \tag{2.18}$$

with μ $(0 < \mu < 1)$ being a learning rate parameter. Now, the LS solution for $\boldsymbol{\Delta w}$ is defined as

$$\boldsymbol{\Delta w} = \left(\boldsymbol{U^H U}\right)^{-1}\boldsymbol{U^H e} \tag{2.19}$$

with $\boldsymbol{e}$ being the $L \times 1$ vector of the residual error, defined as

$$\boldsymbol{e} = \frac{\boldsymbol{y}}{G_0} - \boldsymbol{u} \tag{2.20}$$

where G_0 is the desired linear gain.

2.2.3 DPD Implementation Aspects

Real-Time Predistortion at Forward Path

In a closed-loop adaptive predistorter system like that shown in Fig. 2.11, we can distinguish between the DPD forward path, where the DPD model is applied real-time to the baseband oversampled signal to compensate both for the in-band and out-of-band distortion, and the DPD feedback or adaptation path that takes also the PA baseband oversampled output signal (after proper time alignment and gain compensation) to update the DPD coefficients during training that are being used in the forward path. The DPD training in the feedback path aims at converging to the desired linearization algorithm performance and to adapt to different PA and waveform operating conditions which vary over time. Until the advent of 5G, the speed of such variations was typically several orders of magnitude slower when compared to the forward path time scale (i.e., nanosecond-to-microsecond range) for which, while the digital processing in the forward path is implemented in programmable or hard-wired real-time logic like field-programmable gate arrays (FPGA) or application-specific integrated circuits (ASIC), the feedback path operations are conducted in a processor system doing the LS signal processing and delivering DPD updating speeds in the millisecond-to-second range.

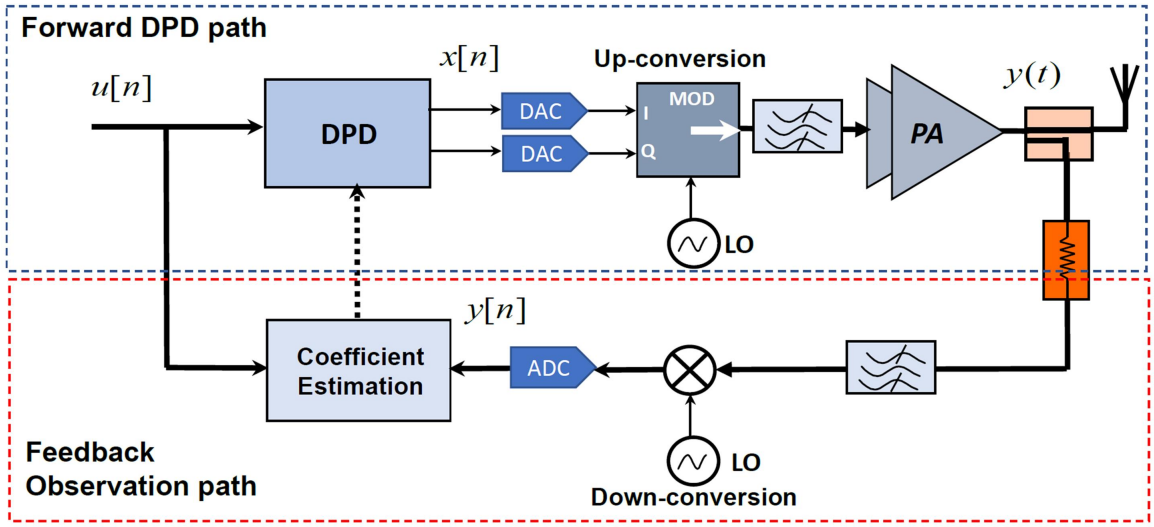


Figure 2.11: Block diagram of a closed-loop adaptive DPD linearization system [Gil20a].

When focusing on the forward path real-time implementation, we can take the distortion signal in (2.14), being $d[n] = \boldsymbol{\phi_u}^T[n]w[n]$. At this stage it is important to consider that in order to make the calculations to compute $\boldsymbol{d}$ different approaches can be followed. One consists in mapping the polynomial model in the programmable logic operating real time, for example by using the Horner's rule as in [Mra12], and therefore implement a given dedicate DPD model. This approach is highly resource consuming but may easily replicate the performance accounted by simulation tools when using the same coefficients. Another approach which is more efficient and features lower processing latency consists in deploying look-up-tables (LUT) that store the values taken by a given polynomial nonlinear function [Gil07b, Mol17, Pha18b]. The input samples are therefore used to index a specific position of the LUT table that gives the output value taken by the function for that given input. In line with the latter approach, the authors in [Ces07] propose combining basic predistortion cells (BPCs) that leverage on the use of LUTs to build the overall DPD. Following an analogous principle, the GMP model in (2.7) can now be applied as

$$\begin{aligned} d\left[n\right] = & \sum_{i=0}^{N_a-1} u_{\mathrm{BB}}\left[n-\tau_i^a\right] g_{i,0}^{(a)}\left(\left|u_{\mathrm{BB}}\left[n-\tau_i^a\right]\right|\right) \\ & + \sum_{i=0}^{N_b-1}\sum_{j=1}^{M_b} u_{\mathrm{BB}}\left[n-\tau_i^b\right] g_{i,j}^{(b)}\left(\left|u_{\mathrm{BB}}\left[n-\tau_i^b-\tau_j^b\right]\right|\right) \\ & + \sum_{i=0}^{N_c-1}\sum_{j=1}^{M_c} u_{\mathrm{BB}}\left[n-\tau_i^c\right] g_{i,j}^{(c)}\left(\left|u_{\mathrm{BB}}\left[n-\tau_i^c+\tau_j^c\right]\right|\right) \end{aligned} \tag{2.21}$$

where the memory depth and cross-memory products are determined by N_a, N_b, N_c and M_b, M_c, while $g_{i,j}^{(k)}(\cdot)$ (with $k = a, b, c$) are the nonlinear functions that depend on the signal envelope products. As explained in [Li21], these nonlinear functions can be described by polynomials

or B-splines, for example, and be mapped to the different LUTs forming the above mentioned BPCs. Now we define

$$g_{i,j}^{(k)}(r) = \sum_{p=0|1}^{P_k-1|0} w_{i,j,p}^{(k)}\, r^p \tag{2.22}$$

where r is the envelope defined as $r = \left|u[n-\tau_i^k \pm \tau_j^k]\right|$, $w_{i,j,p}^{(k)}$ are the coefficients of the polynomial basis functions, P_k is the degree of the polynomial. Each of the basis functions $\varphi_j^u[n]$ (2.17) constitute DPD kernels whose contribution will be weighted by the coefficients. As seen in [Li22], a kernel map $\mathbb{k}$ can be formed with non-repeated kernels of the DPD behavioral model and be indexed by vectors composing i) the delays taken by τ_i^k that are applied to the $\boldsymbol{u}$, ii) the delays taken by $\tau_i^k \pm \tau_j^k$ that are applied to $\left|\boldsymbol{u}\right|$ (envelope term) and iii) the envelope power orders. Considering a model order of O, such map may take the form of

$$\mathbb{k} = (\boldsymbol{\tau^u}, \boldsymbol{\tau^e}, \boldsymbol{p}) = \begin{pmatrix} \tau_1^u & \tau_1^e & p_1 \\ \vdots & \vdots & \vdots \\ \tau_O^u & \tau_O^e & p_O \end{pmatrix} \tag{2.23}$$

and the GMP model be rewritten as

$$d[n] = \sum_{\tau^u,\tau^e,p\in\mathbb{k}} w_{\tau^u,\tau^e,p} u[n-\tau^u] \left|u[n-\tau^e]\right|^p \tag{2.24}$$

Once the kernel map is formed, the implementation scheme that is presented in [Li22], now shown in Fig. 2.12, may be used to efficiently deploy the DPD forward path. In this figure, M_d is the total memory depth accounted as the memory range that need to be allocated to accommodate the samples between the minimum leading and the maximum lagging GMP delay, i.e., $M_d = (-min(\boldsymbol{\tau^u}, \boldsymbol{\tau^e}, 0) + max(\boldsymbol{\tau^u}, \boldsymbol{\tau^e}) + 1)$. The HW will always operate at a system delay or latency of $D = -min(\boldsymbol{\tau^u}, \boldsymbol{\tau^e})$ samples (i.e., minimum leading delay) that will need to be buffered to make the system causal.

When focusing on FPGA implementation, the LUTs that output the power terms will be indexed by the absolute value of the input signal which can be calculated by using a high-precision and high-speed coordinate rotation digital computer (CORDIC) algorithm. The LUTs may be hosted in embedded high-speed grade block random-access memories (BRAMs), and host any kind of precalculated nonlinear function. The resolution of the LUT is determined by its length (number of values stored) and the precision is defined by the bit width used to represent that value. All the LUT-retrieved values are accumulated in shifting memories until the memory depth is filled in for every LUT output. The delay routing blocks in Fig. 2.12 will typically be implemented with flip flops (FFs) to acquire the delayed version of the signals of interest. In between the delay routing blocks, the circular-shaped blocks may be seen as minimum computation units (MCU) that take as input a delayed power term, its corresponding DPD coefficient and the delayed input and computes the product of these values which is added

to the previous MCU. There are as many MCUs as the number of rows of the kernel map. These operations can be handled by FPGA DSP slices. Finally, the predistorted signal is obtained by subtracting the modeled distortion signal to the original signal as per (2.14). Full detail on the

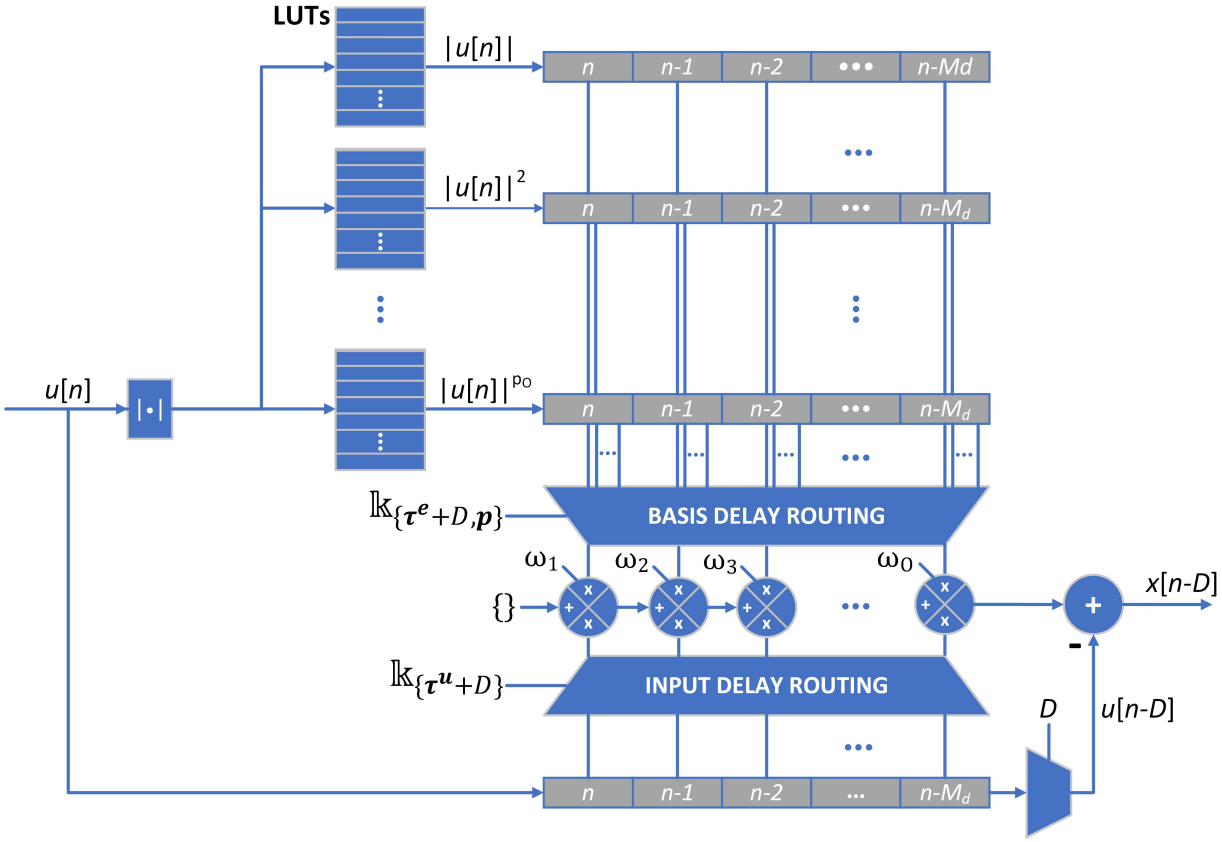


Figure 2.12: DPD forward path GMP LUT-based real-time implementation architecture.

FPGA implementation for this scheme is provided in [Li22] for an advanced 5G ET-PA DPD system running at 614.4 MHz, both considering high-level synthesis (HLS) and register transfer level (RTL) implementation and providing the advantages and pitfalls of each methodology.

Finally, the DPD coefficients $\boldsymbol{w}$ will be updated during training in the DPD feedback path. The numerical methods and calculations behind this process, which are typically computed by an ARM processor embedded in the FPGA, are described in the next section.

Numerical Methods for DPD estimation at Feedback Path

In order to perform DPD coefficient estimation in the feedback path, the LS minimization problem can be solved via Cholesky factorization or normal equations method, QR factorization, Gram-Schmidt methods, or by using the singular value decomposition (SVD). These measurements are introduced as follows:

- **Normal Equations**: This method converts the LS problem to the normal equations first by operating (2.19) as

$$\boldsymbol{U}^H\boldsymbol{U}\boldsymbol{\Delta w} = \boldsymbol{U}^H\boldsymbol{e}. \tag{2.25}$$

The Cholesky factorization can be applied for a full-rank matrix. Being the covariance matrix $\boldsymbol{U}^H\boldsymbol{U}$ positive definite, we can now apply a symmetric form of Gaussian elimination as

$$\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{R}^H\boldsymbol{R} \tag{2.26}$$

where $\boldsymbol{R}$ is an upper-triangular matrix and $\boldsymbol{R}^H$ is lower triangular. Then, what we need to do is solving two following upper-triangular systems

$$\boldsymbol{R}^H\boldsymbol{z} = \boldsymbol{U}^H\boldsymbol{e}, \text{ for } \boldsymbol{z}, \tag{2.27}$$

$$\boldsymbol{R}^H\boldsymbol{\Delta w} = \boldsymbol{z}, \text{ for } \boldsymbol{\Delta w}. \tag{2.28}$$

The operation count for the normal equations method will be $\mathcal{O}(NL^2 + \frac{1}{3}L^3)$. The method is not computational expensive and can run faster when compared to other LS solving solutions. However, its main drawback is that it is sensitive to rounding errors when finite precision arithmetic is deployed to implement the method which may lead to instabilities.

- **QR factorization**: QR is a method also working for full-rank matrices and is among the most commonly approaches used to solve the LS problem. In this method, $\boldsymbol{U}$ is decomposed as

$$\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{R}, \tag{2.29}$$

where $\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_N]$ is an $L \times N$ unitary matrix (i.e., $\boldsymbol{Q}^H\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^H = \boldsymbol{I}$) and $\boldsymbol{R}$ is an $N \times N$ upper-triangular matrix where the diagonal entries r_{jj} are nonzero. Therefore, it can be said that each column of $\boldsymbol{U}$ is a linear combination of the columns in $\boldsymbol{Q}$, that can be expressed as

$$\begin{aligned}
\boldsymbol{\varphi}_{\boldsymbol{1}}^{\boldsymbol{u}} &= r_{11}\boldsymbol{q}_1, \\
\boldsymbol{\varphi}_{\boldsymbol{2}}^{\boldsymbol{u}} &= r_{12}\boldsymbol{q}_1 + r_{22}\boldsymbol{q}_2, \\
&\cdots \\
\boldsymbol{\varphi}_{\boldsymbol{N}}^{\boldsymbol{u}} &= r_{1N}\boldsymbol{q}_1 + r_{2N}\boldsymbol{q}_2 + \cdots + r_{NN}\boldsymbol{q}_N.
\end{aligned} \tag{2.30}$$

From (2.29), the problem in (2.19) has the form of

$$\boldsymbol{Q}\boldsymbol{R}\boldsymbol{\Delta w} = \boldsymbol{e}, \tag{2.31}$$

where the left multiplication by $\boldsymbol{Q}^H$ gives

$$\boldsymbol{R}\boldsymbol{\Delta w} = \boldsymbol{Q}^H \boldsymbol{e}. \tag{2.32}$$

The result is obtained by solving the upper-triangular system in (2.32) for $\boldsymbol{\Delta w}$ with back substitution. The QR approach is more stable than Cholesky and is considered the standard method for solving LS problems (e.g., is used for the MATLAB's backslash operator). The algorithms for computing the QR factorization include the Gram-Schmidt algorithm, Household triangularization, Givens transformation method, Pivoting and Gaussian elimination [Tre97]. In this section, the Gram-Schmidt algorithm and Household triangularization methods for QR factorization are now presented.

Classical Gram-Schmidt algorithm:The classical Gram-Schmidt (CGS) algorithm is a process that generates the orthonormal basis of a set of given linearly independent vectors (in this case, the columns $\boldsymbol{U}$). The Algorithm 1 expresses the CGS. In the j-th step of CGS, the j-th columns of both $\boldsymbol{Q}$ (i.e., the orthonormal basis matrix) and $\boldsymbol{R}$ are generated.

Algorithm 1 Classical Gram-Schmidt algorithm

1: **procedure** CGS ($\boldsymbol{U}$)
2: **for** $j = 1$ to N **do**
3: $\boldsymbol{v}_j = \boldsymbol{\varphi^u}_j$;
4: **for** $i = 1$ to $j - 1$ **do**
5: $r_{ij} = \boldsymbol{q}_i^H \boldsymbol{\varphi^u}_j$;
6: $\boldsymbol{v}_j = \boldsymbol{v}_j - r_{ij}\boldsymbol{q}_i$;
7: **end for**
8: $r_{ij} = ||\boldsymbol{v}_j||_2$;
9: $\boldsymbol{q}_j = \dfrac{\boldsymbol{v}_j}{\boldsymbol{\varphi^u}_j}$;
10: **end for**
11: **end procedure**

The disadvantage of the CGS algorithm is its sensitivity to the cumulative effects of rounding errors at the processing system. To overcome this problem, the modified Gram-Schmidt (MGS) algorithm is proposed. In CGS, the lengths of the orthogonal projections of $\boldsymbol{v}_j = \boldsymbol{\varphi^u}_j$ onto $\boldsymbol{q}_i$, $i = 1, ..., j-1$ are computed and then these projections are subtracted from $\boldsymbol{v}_j$. Differently, in MGS the length of the projection of $\boldsymbol{v}_j$ onto each q_i, $i = 1, ..., j-1$ is computed and then subtracted from $\boldsymbol{v}_j$ as soon as the q_i is known.

Algorithm 2 Modified Gram-Schmidt algorithm

procedure MGS ($\boldsymbol{U}$)
  for $i = 1$ to N **do**
    $\boldsymbol{v}_i = \boldsymbol{\varphi^u}_j$;
  end for
  for $i = 1$ to N **do**
    $r_{ii} = ||\boldsymbol{v}_i||_2$;
    $\boldsymbol{q}_i = \dfrac{\boldsymbol{v}_i}{r_{ii}}$;
    for $j = i + 1$ to N **do**
      $r_{ij} = \boldsymbol{q}_i^H \boldsymbol{v}_j$;
      $\boldsymbol{v}_j = \boldsymbol{v}_j - r_{ij}\boldsymbol{q}_i$;
    end for
  end for
end procedure

In other words, while in CGS we take each vector, one at a time, and make it orthogonal to all the previous vectors, in MGS we take each vector and modify all the forthcoming ones to be orthogonal to it. Despite the two variants mathematically equivalent, MGS is more immune to rounding error instability than CGS and features $\mathcal{O}(2LN^2)$ operation count.

Householder triangularization: This is another well-known approach to calculate the QR factorization. The stability of this method is equivalent to that of MGS but features slightly lower computational cost. The Householder algorithm is a process of triangularizing a matrix by a sequence of unitary matrix operations. The essence of this method is at the i-th step, doing a left-multiplication $\boldsymbol{U}$ by a matrix $\boldsymbol{Q}_i$ so that the entries from the $i+1$-th one of the i-th column of $\boldsymbol{U}$ become zeros. The output of the algorithm are the upper triangular matrix $\boldsymbol{R}$, and the vectors $\boldsymbol{v}$ which can be used to form the matrix $\boldsymbol{Q}$ so that $\boldsymbol{U} = \boldsymbol{QR}$. In the algorithm, the $\boldsymbol{U}_{i:i',j:j'}$ is defined as the $(i' - i + 1) \times (j' - j + 1)$ submatrix of $\boldsymbol{U}$ with the upper left corner $\boldsymbol{\varphi^u}_{ij}$ and the lower right corner $\boldsymbol{\varphi^u}_{i'j'}$. In case of presenting the subvector of a single row or column, we write $\boldsymbol{U}_{i,j:j'}$ or $\boldsymbol{U}_{i:i',j}$. The operation count for the algorithm is of $\mathcal{O}(2NL^2 + \frac{2}{3}L^3)$.

Algorithm 3 Householder algorithm

procedure HHA ($\boldsymbol{U}$)
  for $i = 1$ to N **do**
    $\boldsymbol{\phi_u} = \boldsymbol{U}_{i:N,i}$;
    $\boldsymbol{v}_i = sign(\boldsymbol{\phi_u}_1)||\boldsymbol{\phi_u}||_2\boldsymbol{e}_1 + \boldsymbol{\phi_u}$;
    $\boldsymbol{v}_i = \boldsymbol{v}_i||\boldsymbol{v}_i||$;
    $\boldsymbol{U}_{i:L,i:N} = \boldsymbol{U}_{i:L,i:N} - 2\boldsymbol{v}_i(\boldsymbol{v}_i^H \boldsymbol{U}_{i:L,i:N})$;
  end for
end procedure

- **SVD**: SVD is another matrix factorization method for either real or complex matrices. This LS solving method can be applied when the matrix $\boldsymbol{U}$ is close to rank-deficient since it has very good stability properties. The decomposition of $\boldsymbol{U}$ is formulated as

$$\boldsymbol{U} = \boldsymbol{V}\Sigma\boldsymbol{W}^H \tag{2.33}$$

where each column of the $N \times N$ matrix $\boldsymbol{W}$ (or *right singular vector*, please note that the variable naming has nothing to do with neither DPD coefficient vectors nor DPD coefficient increment vectors) is the unit vector of the unit sphere $\boldsymbol{S}$ (considering the 2-norm unit sphere). Each column of the $L \times N$ matrix $\boldsymbol{V}$ (or *left singular vector*) is the unit vector oriented in the direction of a principle semiaxis of $\boldsymbol{US}$ and, therefore, the matrices $\boldsymbol{V}$ and $\boldsymbol{W}$ are unitary. Finally, Σ is an $N \times N$ diagonal matrix with positive entries. When full SVD is performed, $\boldsymbol{V}$ is $L \times L$, $\boldsymbol{W}$ is $N \times N$ and Σ is $L \times N$.

The LS problem in (2.19) is now seen as

$$\boldsymbol{V}\Sigma\boldsymbol{W}^H\boldsymbol{\Delta w} = \boldsymbol{e}. \tag{2.34}$$

Applying left multiplication by $\boldsymbol{V}^H$ we have that

$$\Sigma\boldsymbol{W}^H\boldsymbol{\Delta w} = \boldsymbol{V}^H\boldsymbol{e}, \tag{2.35}$$

and by setting $\boldsymbol{z} = \boldsymbol{W}^H\boldsymbol{\Delta w}$, (2.35) becomes

$$\Sigma\boldsymbol{z} = \boldsymbol{V}^H\boldsymbol{e}. \tag{2.36}$$

By solving the diagonal system in (2.36) for $\boldsymbol{z}$, then we compute $\boldsymbol{\Delta w} = \boldsymbol{W}\boldsymbol{z}$. The operation count for the algorithm is typically $\mathcal{O}(2NL^2 + 11L^3)$. The SVD can be solved via eigenvalue decomposition, Golub-Kahan bidiagonalization or Lawson–Hanson–Chan (LHC) bidiagonalization [Tre97].

2.3 Crest Factor Reduction Techniques

CFR techniques are aimed at reducing the peaks of the modulated waveform to a satisfactory level. Due to the potentially added in-band and out-of-band distortion, CFR does not necessarily introduce system gain; but, thanks to the reduction in the peaks, it is possible to operate the PA at higher average power and therefore closer to its saturation point, where it is most efficient. As depicted in Fig. 2.10, CFR is usually combined with DPD. The positive impact on the radio equipment when combining these two digital linearization techniques has been detailed in Section 1.1.2.

The PAPR reduction techniques may be divided in three groups according to the taxonomy shown in [Rah13]:

- In the first group, the signal distortion techniques reduce the PAPR by distorting the OFDM signal before the PA at the expense of some tolerable increase in the in-band and out-of-band distortion. Clipping techniques may be classified into hard-clipping, soft-clipping and companding. In hard-clipping the output signal is strictly limited at the established threshold, while in soft-clipping the output signal follows a piecewise law where several threshold levels are defined. In the companding technique the dynamic range of the signal is compressed at the transmitter side by means of a memoryless transformation (i.e., companding function). These clipping techniques require some kind of spectral shaping procedure to mitigate the clipping noise that appears as spectral regrowth in the adjacent channels. To cope with the spectral shaping of the clipping noise, several techniques have been proposed, such as clipping pulses [Kim07a], pulse windowing [Vaa05] or noise shaping [Sau04].

- The second group includes the multiple signaling and probabilistic techniques that either modify the OFDM waveform (i.e. by introducing phase shifts, adding peak reduction sub-carriers or changing the symbol constellation points) or generate multiple permutations of the OFDM signal and transmit the one with the minimum PAPR. In this group we find the selective mapping (SLM) [Bäu96], partial transmit sequences (PTS) [Mül97], tone injection and tone reservation [TM99], active constellation extension (ACE) [Kro03] and the interleaved OFDM [Hil00, Han05], among others.

- In the third group some coding schemes like linear block coding, turbo coding or Golay Complementary Sequences are provided both for error detection and correction and to perform PAPR reduction. The idea of the coding schemes is to reduce the occurrence probability of the same phase of the signals by selecting the codewords that minimize the PAPR (avoiding in-phase addition of signals) in the transmission. Several coding techniques have been published in literature, such as for example the Simple Block Coding [Fra98], Complement Block Coding [Jia05] or the Modified Complement Block Coding [Jia04], among others.

The CFR solutions are typically analyzed in terms of implementation complexity, bit rate loss, induced distortion and in terms of power increase. The signal distortion techniques are good solutions in terms of implementation complexity, preserving the bit rate, and not requiring a power increase but, on the contrary, the EVM and BER may increase. The multiple signaling and probabilistic techniques are of higher complexity but do not provoke distortion. The bit rate loss and power increase will depend on the specific technique. For instance, while in ACE the symbol constellation modification requires higher power but does not imply bit rate loss, PTS does not require power increase but does imply a bit rate loss due to the need of side information (i.e. sent to inform the receiver of what modifications have been implemented at the transmitter).

In this thesis, the CFR technique used to reduce the PAPR of the modulated signals is the peak cancellation (PC) scheme described in [Kim07b], due to the reasonable implementation complexity and tuning capabilities. As shown in the block diagram in Fig. 2.13, it is based in the clipping and filtering techniques:

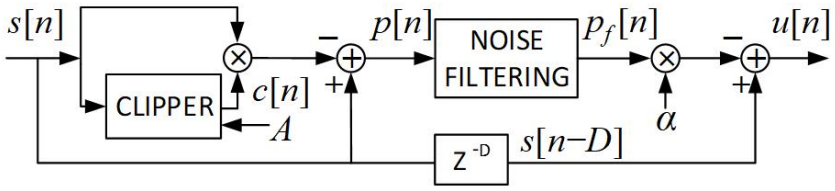


Figure 2.13: Block diagram of the PC-CFR technique.

In a first stage, a peak signal $p[n]$ is generated as

$$p[n] = s[n] - s[n]c[n] \tag{2.37}$$

being $s[n]$ the input signal and $c[n]$ the clipping signal defined as

$$c[n] = \begin{cases} \frac{A}{|s[n]|} if |s[n]| > A \\ 1 \quad \text{if } |_s[n]| \leq A \end{cases} \tag{2.38}$$

with A being a clipping threshold. In a second stage, the peak signal $p[n]$ is filtered (noise shaping), obtaining $p_f[n]$. Finally, the reduced PAPR output signal is calculated by subtracting a weighted version of the filtered peak signal from the original input signal (after proper time alignment),

$$u[n] = s[n - D] - \alpha p_f[n] \tag{2.39}$$

with α being the subtraction parameter. In order to guarantee the PAPR reduction factor for a given set of threshold and subtraction parameters, the output signal is fed back to the input and the process is iterated four times more. By benchmarking in advance the PAPR reduction factor and the quality of the signal for all the different combinations of threshold and subtraction parameters, it is therefore possible to tune the PC-CFR trading-off between induced distortion and PAPR reduction to meet the requirements imposed by the application scenario.

Finally, one must be aware that by using the PC-CFR technique as stand-alone digital linearization scheme may not suffice to provide any relevant benefit to the system or meet the required specification. The rationale to include it in the digital linarization system, however, are listed as follows: i) The predistorter may be assisted by the CFR block as a protection mechanism to minimize the effects of having signals with larger PAPR than those with the highest probability of occurrence and that are typically used for training the DPD or calculating a specific set of coefficients in non-adaptive DPD architectures. The low probability of occurrence

but highest PAPR transmissions can feature, for example,2-to-3 dB higher PAPR than the mean statistical value. ii) Preventing the peaks from going hard into compression allows keeping under control the ACPR values and thus increasing the output power to have further system gain and power efficiency. In adaptive DPD architectures, the CFR mechanism facilitates compensating the response of a highly compressed PA (i.e. by the predistorter) by easing the parameter identification and the convergence of the DPD engine. It helps to prevent excessive PAPR values of the predistorted signal being produced (always some dBs higher than the nominal ones as the compressed AM/AM curve needs to be inverted) which otherwise could affect the quantization performance at DAC devices or impose hard restrictions on the mean power at the output of such devices to avoid DAC saturation.

Combined CFR and DPD Use Case

Trading off Efficiency and Spectral Coexistence in Emergency Roll-Out Radio

This use case summarizes a highly detailed book chapter contribution in [LB20]. It highlights the benefits of employing digital linearization in a not so common application scenario, where a cognitive and agile broadband public protection and disaster relief (BB-PPDR) radio may need to be deployed to replace compromised radio access and backhaul equipment or extending coverage with temporary infrastructure equipment, and coexist with other wireless communication or broadcasting technologies such as LTE or DVB-T. In the case of a cross-border problematic, a rapid context-aware on-field assessment of the situation might also be necessary (i.e. scan the spectrum of interest to identify interferers and jammers), in order to apply a coarse grain minimization of interference sources. The BB-PPDR radio transmission must satisfy the following key performance indicators (KPI):

- A high degree of flexibility of the transmitted waveform, able to be reconfigured according to the local spectrum landscape and coexist with adjacent or even in-band secondary or co-primary transmissions (living in in-band spectral holes) [FB15].
- Given the previous requirement, the transmitted BB-PPDR waveform has to satisfy a spectral mask where the out-of-band (OoB) emissions are minimized.
- BB-PPDR also require a high spectral efficiency considering for example the on-site operational needs for HD video transmission.
- The high peak-to-average power ratio (PAPR) of the wideband multicarrier waveforms used in BB-PPDR limits the transmitter energy efficiency. The PA needs to be operated with significant back-off avoid the waveform peaks falling in the non-linear region and to meet the signal quality requirements. Such operation is extremely inefficient and not convenient in an energy-constrained scenario (i.e. without electrical grid availability). The

use of digital linearization techniques contributes to optimally handle the efficiency vs linearity trade-off.

Special focus is therefore given on the last KPI related to the energy-efficient linearization of the transmitters, which is critical for guarantying low OoB emissions, exploiting the inherent spectral efficiency of the proposed transmitted waveform and avoid impairing the performance of flexible waveforms. This use case leverages on the use of an OFDM alternative for the communication system based on the filter-bank multi-carrier (FBMC) modulation that was one of the candidate waveforms proposed initially for the 5G communication systems [Gue17], due to their superior spectral efficiency and spectral shape characteristics (resulting in nearly optimal OoB emissions). This waveform has been widely used to benchmark the main contributions of the thesis.

Digital linearization architecture and test setup

In this work, we have considered the high-level digital linearization architecture shown in Fig. 2.14, which is based on combining a 5-stage PC-CFR and a GMP direct-learning DPD. To model the non-linearities of the PA, the test bench in Fig. 2.15 is used to get the PA output waveforms that will be compared to the originally sent ones after time alignment and gain compensation.

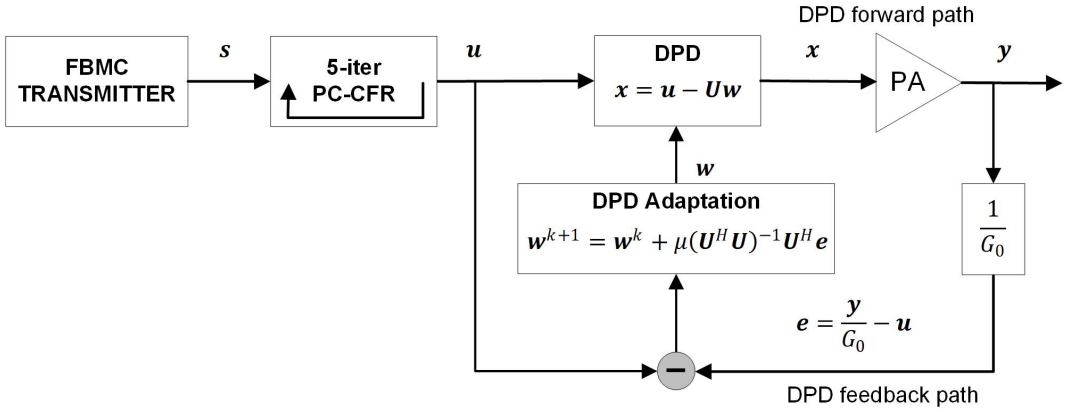


Figure 2.14: CFR and direct learning DPD architecture proposed for the BB-PPDR BS.

The chosen configuration for the GMP model in (2.7) leads to about 100 DPD coefficients which is a considerable amount but also a reasonable figure considering a PA showing very strong nonlinearities due to operating under hard compression (no CFR is applied yet), and that this number can be then reduced significantly by using a feature selection technique (like that proposed in Chapter 3). By considering the GMP structure in the predistorter, the coefficients will be estimated iteratively according to the method shown in Fig. 2.14.

In the experiments that were conducted, the FBMC system was validated considering LTE-like requirements, as specified in [TSIE14], regarding the ACPR and the EVM. The ACPR

requirement is set at -45 dBc and the EVM requirement is set at 18.5 per cent, 13.5 per cent, 9 per cent and 4.5 per cent considering the QPSK, 16-QAM, 64-QAM and 256-QAM subcarrier modulations, respectively. The test bench that was used to tune, fast prototype and evaluate the digital linearization techniques, employs 20 MHz bandwidth downlink signals which are modulated with fast convolution (FC) FBMC and that follow the structure described in [FB17]. In the tests, 64-QAM modulated FBMC signals were employed. These waveforms feature one-by-one subcarrier deactivation to optimise the spectrum occupation and allow in-band spectral coexistence with co-primary systems operating. Two spectrum holes were created through subcarrier deactivation at ±4.5 MHz (from center frequency) to allow coexistence with two LTE signals with 5 MHz and 1.4 MHz bandwidth, respectively.

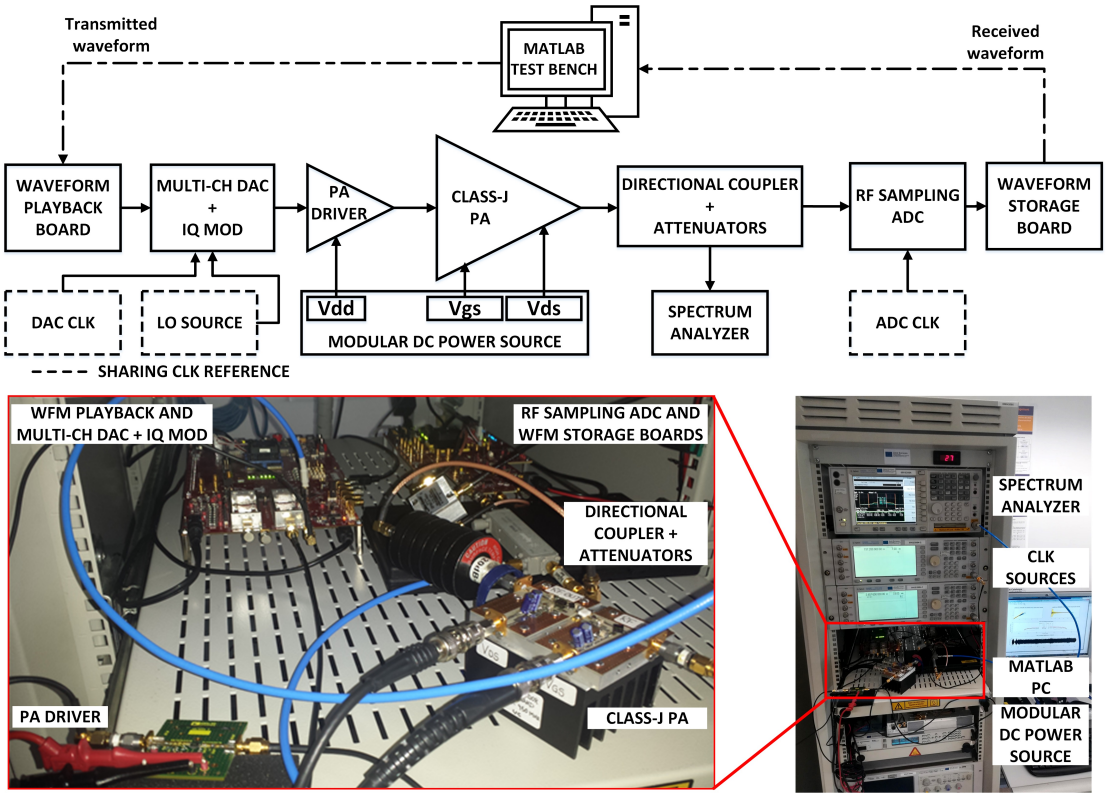


Figure 2.15: Digital linearization test bench.

The FBMC signals followed a frame and timing structure similar to that used in LTE (LTE FDD Rel. 9) for which the original baseband sample rate for 20 MHz bandwidth is 30.72 MHz and the signals are then oversampled by five at 153.6 MHz (maximum DPD signal bandwidth) to allow for OoB compensation when predistorting. These waveforms featured around 11.5-dB PAPR at baseband. The MATLAB-controlled digital linearization test bench employed some commercial boards for waveform playback and data capture, digital-to-analogue conversion (DAC), in-phase and quadrature-phase (IQ) modulation, and analogue-to-digital conversion (ADC) for direct RF sampling. The DPD signal is in these experiments IQ modulated to the RF frequency

to feed a class-J PA based on a GaN high-electron-mobility transistor (HEMT) biased at 28 V drain-to-source voltage and -3.3 V gate-to-source voltage. The RF operating frequency is located at 875 MHz which is near the center of the 3GPP band 26 (downlink) also employed for PPDR. The waveform at the PA output is then attenuated and RF sampled at 2457.6 MSa/s, to be further decimated and time aligned with the signal that was originally sent.

Eight iterations (Tx-Rx and DPD processing loops) are used to reach the results found for different output power experiments (considering no DPD, DPD and CFR+DPD) that are shown in Table 2.1. At every experiment, the worst case IMD5, IMD3 and in-band channel power rejection (IBPR) values are shown together with the EVM and the PA drain efficiency. Fig. 2.16 shows the spectra plots and constellation before and after linearization for the +33 dBm (top) and +35 dBm (bottom) output power levels and the corresponding symbol constellations. The results show the benefits of using the digital linearization techniques. If the same PA is kept and only back-off was considered to meet the ACPR requirements, the mean output power would be reduced by more than 13 dB and the PA would be operated very inefficiently.

Table 2.1: Digital linearization experimental results at different power levels.

Case	**Pout [dBm]**	**Worst IMD5 [dBc]**	**Worst IMD3 [dBc]**	**Worst IBPR [dBc]**	**EVM [%]**	**Drain Eff. [%]**
No DPD	20	-51.85	-45.74	-42.44	6.75	4
No DPD	33	-45.67	-30.83	-26.36	8.55	35
DPD	33	-55.57	-48.01	-44.98	1.12	35
No DPD	34	-46.72	-29.48	-24.38	11.41	41
DPD	34	-53.28	-43.16	-42.38	1.30	41
No DPD	35	-46.28	-28.35	-22.76	12.32	47
CFR[a]+DPD	35	-53.76	-45.06	-35.26	3.10	47
CFR[b]+DPD	35	-56.11	-53.13	-21.70	5.27	47

[a] 2.6 dB PAPR reduction
[b] 4.1 dB PAPR reduction

When considering a nominal operating output power of +33 dBm, the IMD3 and the IBPR can be improved by about 18 dB and the EVM can be reduced more than 7 percent points, both on account of the DPD. When the output power is increased to +34 dBm, when only applying DPD the IMD3 (or ACPR) requirement of -45 dBc is no longer met since the predistorter cannot compensate the PA hard compression when inverting its response. If CFR is applied before DPD, the results show how that is feasible to reach at least a 2 dB higher output power level, while meeting the ACPR and EVM requirements, by clipping the signal with some controlled in-band distortion. Thanks to increasing the output power, when comparing back-off operation with the nominal one at +33 dBm when applying DPD, the drain efficiency is enhanced by 31 percent points. As a result of applying CFR+DPD the drain efficiency can be enhanced by at least 12 percent points more at the cost of increasing the in-band distortion. For this reason, the BS

needs to be aware of the presence of co-primary signals and the interference received by the UEs to trade-off between the FBMC transmitter efficiency and the allowable interference to the co-primary system. A more detailed analysis in this regard is provided in the next section but the results indicate that with CFR+DPD the PA drain efficiency could overpass the 50 per cent mark; The PA output power could be increased beyond 36 dBm if the interference to the in-band channels requirement was alleviated or the in-band co-primary channels were freed (in which case the FBMC waveform could also occupy the whole 20 MHz channel).

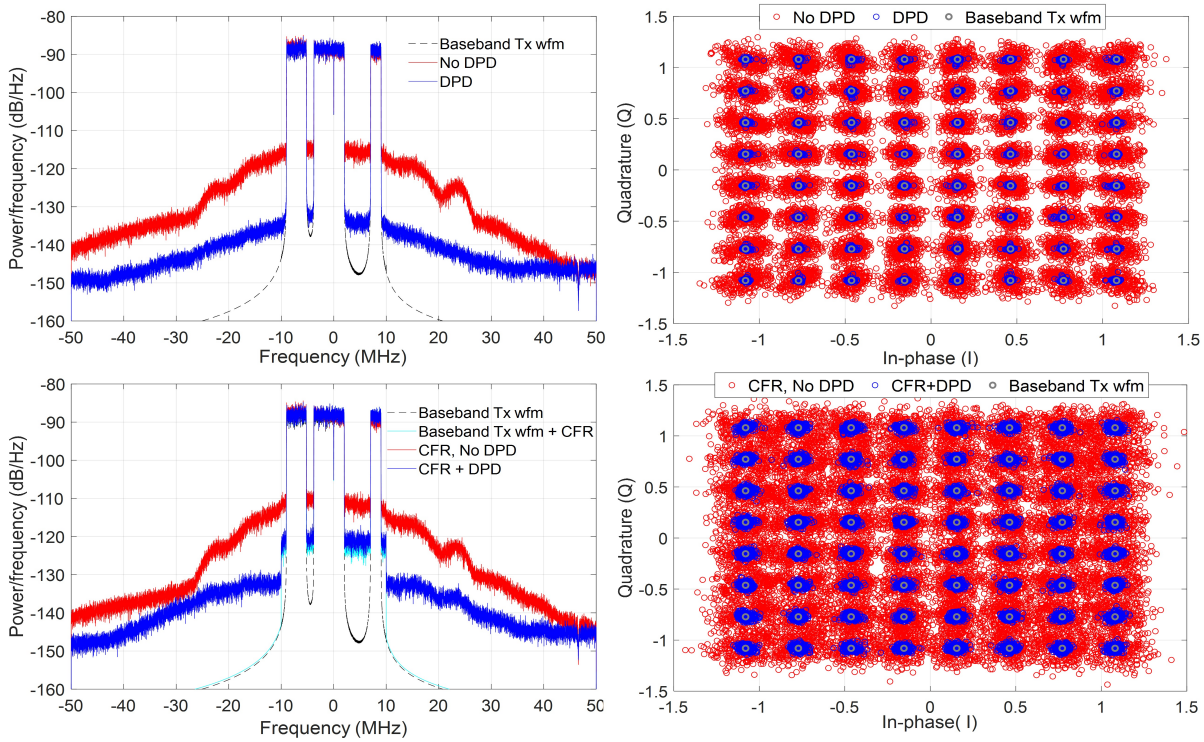


Figure 2.16: Spectra plots and constellations of the linearization experiments at +33 dBm (top) and +35 dBm (bottom) mean output power levels.

Impact of the digital linearization in spectral coexistence

Now we emulate a coexistence scenario where the 1.4 MHz bandwidth and 5 MHz bandwidth LTE signals that coexist with the FBMC BB-PPDR transmission are received by different users. These users will filter out the content of the FBMC signal but they will receive a certain amount of FBMC to LTE interference in-band. The amount of interference is induced in these experiments by sweeping the power difference between the LTE signals (i.e. appearing in the deactivated subcarriers) and the broadband FBMC one from 0 dB to -25 dB. Fig. 2.17 (top) illustrates this concept for a power difference swept between the LTE signals (which are set at same output power level) and the FBMC one. In this evaluation, we have considered the FBMC system experimental data taken with and without PA digital linearization to demonstrate its

beneficial impact from the waveform coexistence point of view. The two LTE signals are 256-QAM modulated and the EVM is evaluated under the different LTE vs FBMC power ratios. As expected, the results for the 5 MHz bandwidth signal will be worse than those of the 1.4 MHz bandwidth signal since the in-band noise level will be higher in the first case. In Fig. 2.17 (bottom), the benefits of using digital linearization towards spectral coexistence are demonstrated. If we observe the results for the 5 MHz bandwidth LTE signal (worst case EVM), when comparing the traces with and without DPD and the LTE-FBMC power difference at which the 4.5 per cent EVM threshold is met, about 18.5 dB of SINR are gained. Near 6 dB will be gained when using CFR and DPD (i.e. for a 2.6 dB PAPR reduction setting). The 12.5 dB degradation in terms of robustness against interference when applying CFR and DPD (compared to applying only DPD) is coherent with the data shown in Table 2.1.

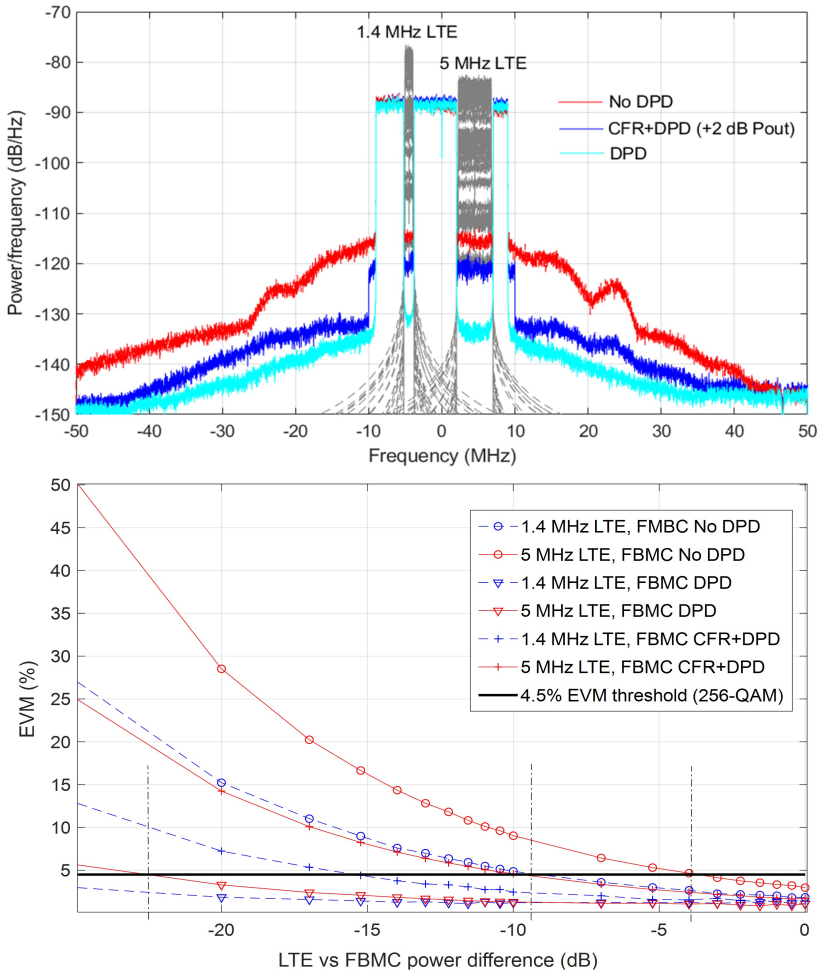


Figure 2.17: Waveform coexistence spectra plots (top) and EVM (bottom) for different LTE vs FBMC power ratios and digital linearization experiments.

In this example, applying CFR may contribute to increase the FBMC transmitter output power and PA drain efficiency by 2 dB and 12 percent points, respectively. Moreover, in a

coordinated scenario the amount of FBMC in-band induced CFR distortion could be adaptively controlled according to the LTE user's SINR information being shared with the FBMC BS. In such scenario, we would consider the LTE-compatible frame structure of the FBMC system physical layer and assume that the rest of the protocol stack remains LTE both at the BS and the UE sides of the BB-PPDR radio access network. By adapting the PAPR reduction value accordingly, it is possible to meet both the FBMC EVM and ACPR requirements and tune the in-band distortion level to a level that maximizes the FBMC transmitter efficiency and at the same time allows the LTE signals meeting the most stringent EVM requirement at 256-QAM subcarrier modulation.

Chapter 3

Machine Learning Techniques for Polynomial SISO Digital Predistorters

3.1 Dimensionality Reduction and Machine Learning for DPD: Early Works

3.1.1 Introduction to DPD Computational Complexity Reduction

As introduced in Section 1.1 and Section 2.2.3, the DPD adds complexity to the transmitter analog and digital signal processing stages to handle the linearity versus efficiency trade-off. In this chapter, we take advantage of machine learning techniques (ML) to reduce i) the length of the DPD training signals to be processed in the feedback path to find or update the coefficients and then ii) to reduce also the number of features (or basis functions, components, variables, dimensions) in the DPD model in the forward and feedback paths, as shown in Fig. 3.1 for the DPD modeling matrix $\boldsymbol{U}$.

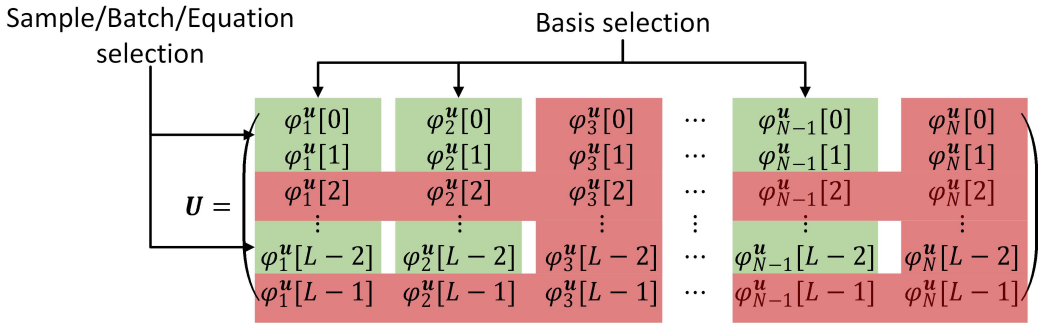


Figure 3.1: Training data and basis selection concepts applied to reducing the length and dimensionality of a behavioral modeling matrix.

The goal is to save as many hardware logic resources and memory as possible (i.e., DSP slices, configurable logic blocks, FFs, configurable logic block (CLB) LUTs, BRAM), speed-up the time to train or update the DPD coefficients, and guarantee robust and well-conditioned identification. When handling complex nonlinear wideband multi-antenna systems, every gain in this regard may be very important. We can distinguish between feature selection techniques, that are applied in the DPD forward path and take the most relevant variables from a random set of original variables, and feature extraction techniques that transform the original set of variables and create a new reduced set with variables that are linear or nonlinear combinations of the original ones. In the feedback path the training data may be selected through either sample, batch or equation selection depending on whether the selection is made taking sparse samples (when operating directly with PA input and output waveforms, before building $\boldsymbol{U}$), a batch or several batches of data (also considering similar operation), or equations of the DPD modeling matrix (the selection is made by operating with the whole U matrix).

A classification and description of feature selection techniques (filter, wrapper and embedded methods) and feature extraction techniques (linear and nonlinear dimension reduction methods) is provided in [Gil20b]. These techniques can also be properly combined by doing an *a priori* off-line search to reduce the number of basis functions of the forward path DPD function (feature selection) and then applying a feature extraction technique in the feedback path [Pha18c].

In the following, several sample selection, feature selection and feature extraction techniques that have been used for DPD dimensionality reduction are detailed together with specific application use cases that were deployed in the early research works of the thesis.

3.1.2 Sample Selection for Training Data Length Reduction and DPD robustness

In the DPD feedback path, in charge of the estimation and the update of the coefficients, the iterative DPD training must be made with sufficient waveform statistical representativity in terms of power dynamics. This frequently translates into having to process a huge amount of samples from different training waveforms covering the PAPR statistical distribution for a given set of conditions (bandwidth, modulation and subcarrier modulation), since the optimality of the extracted DPD coefficients using LS strongly depends on the number of samples used in the computation. The LS estimator may need *a priori* a large number of data samples (or equations) to obtain the best approximation of the coefficients. From a computational complexity point of view, we want the number of data samples be as small as possible. However, using short sequences of data samples leads to [Gua12]: a) the ill-conditioning problem due to the rank deficiency of LS matrices; and b) the statistical mismatch problem, because the short data sequence often cannot fully represent (accuracy problem) the statistical property of the input domain.

As noted in Section 2.3, by using CFR one can better condition the DPD training by suppressing the lowest occurrence probability waveforms with highest PAPR and optimize its dynamic range (i.e., by limiting the predistorted signal PAPR expansion required to compensate for the PA behavior at saturation). This may also help to reduce the number of DPD training iterations but, even within a given training iteration, the input-output waveform samples used in the DPD feedback path are critical in order to capture the PA nonlinearities and memory effects. Not only the length of the training set is important. If these samples are chosen randomly (either in the form of contiguous sample batches or sample by sample), the DPD coefficients will not be valid to cover the full set of nonlinear effects for a waveform configuration and PA operating condition and the system will be ill conditioned. Thanks to conveniently capturing representative training data samples, the sample selection methods (SSM) can reduce the number of DPD training operations and memory requirements, and speed up the overall DPD adaptation time or even reduce the ADC sample rate requirements in the DPD observation path. In addition, these methods can also contribute to uncorrelate the observation errors in adjacent samples to get better performance.

Several SSMs have been reported in literature such as undersampling, sample selection upon QR decomposition with column pivoting [Due17], gradient sampling or the histogram-based methods. An comprehensive overview on these techniques and its application to DPD is found in [Kra20]. The most commonly employed are summarized in the following:

- **Undersampling**: This scheme [Wan17b] relies on the fact that in order to estimate the model coefficients it is not necessary to capture consecutive samples of the PA output signal $\boldsymbol{y}$ if $L_{us} >> N$, being L_{us} the length of the undersampled signal. That is, the reduced undersampled set must be significantly larger (i.e., more than two orders of magnitude larger) than the model order (number of basis functions and of DPD coefficients) to minimize the measurement noise effects. To undersample, 1 out of every RF_l samples are selected, where $RF_l \in \mathbb{N}$ is the undersampling or length reduction factor (i.e., in the complexity reduction context we will refer to the reduction factor as RF, and the l subscript will imply that the reduction is applied to the length), and therefore $L_{us} = L/RF_l$. This approach is typically used in moderately nonlinear scenarios since, despite taking non-consecutive samples helps to uncorrelate eventual errors, when the length reduction factor or undersampling factor RF_l is increased, there is higher probability of not capturing relevant information, when compared to other SSMs. This is because, typically, the PA output waveform nonlinearities are unevenly distributed all over the waveform.

- **Gradient Sampling**: In order to reduce the computation cost to solve least squares, an adaptive importance gradient sampling method is proposed in [Zhu16] to handle large sample size. This technique is applied to a direct learning DPD architecture in [Kra20]. Given the basis matrix $\boldsymbol{U}$ with $L \times N$ dimension, and all the samples of the measured

distortion signal $\boldsymbol{d} = \boldsymbol{y} - \boldsymbol{u}$, the gradient values $\nabla[n]$ for all the rows in matrix $\boldsymbol{U}$ are calculated (for a given DPD training iteration) as

$$\nabla[n] = \left\| \boldsymbol{\phi_u}^T[n] \left(d[n] - \boldsymbol{\phi_u}^T[n] \boldsymbol{w}_0 \right) \right\| \tag{3.1}$$

where $\|.\|$ is the l_2-norm, $\boldsymbol{\phi_u}^T[n]$ is the $n-th$ row of matrix $\boldsymbol{U}$, $d[n]$ is the $n-th$ element of vector $\boldsymbol{d}$ and $\boldsymbol{w}_0$ is considered an initial good guess for solving the coefficient increment vector (as per equation 2.19), being a zero vector at the first training iteration, and the calculated coefficients of the previous DPD training iteration otherwise. Then, the probability that the n-th row of matrix $\boldsymbol{U}$ is taken into account in the DPD adaptation is found as

$$p_n = \frac{L_{gs} \nabla[n]}{\sum_{n=1}^{L} \nabla[n]}, \tag{3.2}$$

where the reduced set of selected samples $L_{gs} = L/RF_l$ to be considered for the update of new DPD coefficients will be built based on these probabilities (those with higher probability will be considered first).

- **Histogram based**: These techniques are based on filling in with a specific number of samples the bins of a histogram which covers the DPD oversampled baseband transmitted signal amplitude statistics, to obtain a statistically representative DPD training set. We can differentiate between methods considering the statistics of $\boldsymbol{u}$ and others targeting the predistorted signal $\boldsymbol{x}$ which is present at the input of the PA and that will include *a priori* knowledge of the PA characteristics. The selected samples may be allocated in the histogram bins either randomly or sequentially processing the afore mentioned waveforms. To ensure the independence of selected samples, a minimum number of sample time distance can be set [Kra20]. Regarding the distribution of samples within the histogram bins, two options are typically considered. The first consists of evenly distributing the samples in the histogram bins (all the bins will contain the same number of samples) to cover the whole AM/AM characteristics of the PA. Such distribution is not optimized in terms of providing more information for the points involving high probability of occurrence and higher nonlinearities which are more important for PA/DPD modeling, since all the regions are equally covered. To overcome such limitation, the second approach consists of accounting both for the AM/AM characteristics of the PA and the statistics of the transmitted signal, to select a variable number of samples at each bins. By following this strategy, the regions with strong nonlinearity and/or high probability will contain more histogram bin samples than those with lower probability of occurrence and more linear behavior. One example of the latter is the genetically optimized histogram (GOH) which is formulated for histogram sample selection in [Kra17]. Optimizing the histogram with

GOH may consume significant computational cost and time for which if the PA operation, or the transmitted waveform, vary significantly over time and such variations may be produced rapidly, this approach may be unfeasible unless GOH is applied in advance for a set of known cases. Alternatively, the evenly distributed histogram sample selection may provide similar performance by increasing the number of samples per bin (or decreasing the training data length reduction factor RF_l.

These two methods are depicted in Fig. 3.2 where an evenly distributed histogram is filled with two samples per bin, and according to the AM/AM $\boldsymbol{x} - \boldsymbol{y}$ characteristic of the PA and the probability distribution function (pdf) of the $\boldsymbol{u}$ signal AM, a GOH algorithm conveniently allocates a variable number of samples at each bin.

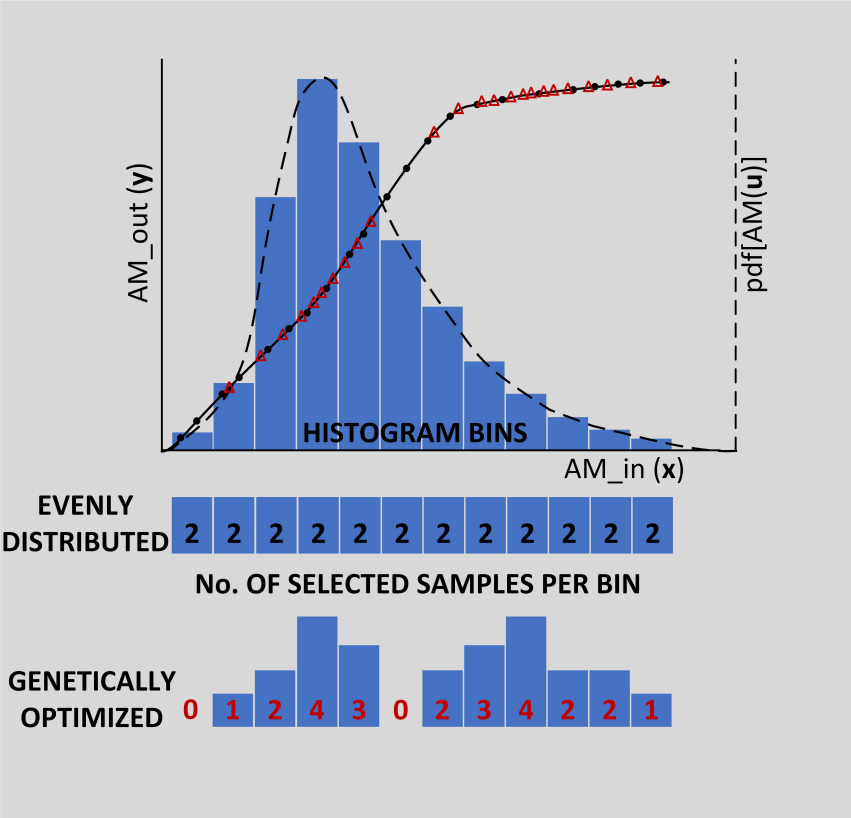


Figure 3.2: Histogram-based SSM concept for evenly distributed and genetically optimized sample capture.

Sample Selection Use Case

Enhancing GaN PA DPD robustness under burst-like 5G waveforms

This use case deploys DPD strategies to linearize a GaN PA scenario fitting 4G-5G radio base stations that operate with high-PAPR OFDM-based burst-like waveforms. In particular, such DPD strategies are combined with a 2D histogram-based SSM in order to guarantee robust performance. Such scenario was proposed in the frame of the 2015 PA Linearization through

Digital Predistortion Student Design Competition (SDC-8) [Shi16, LB16]. The PA, that was operated with 40-MHz bandwidth signals (160 MHz accounting for the total DPD bandwidth), was showing substantial memory effects due to trapping phenomena [Teh12] being provoked by the gate on-off switching.

First, a statistical characterization block giving information about the waveform PAPR distribution and storing multiple test signals for the predefined PAPR ranges of interest, has been created. The test signals showing a PAPR around 13.5 dB were the ones with the highest statistical significance. Once desirable test signals were made available, an initial characterization of the PA through the measurement setup was done by observing the AM-AM and AM-PM characteristics, the spectral shape (see Fig. 3.3) and measurement parameters like the NMSE and the ACPR. To evaluate this data, proper time alignment between the transmit and receive waveforms is required. To find the fractional time delay between them, the two waveforms were upsampled by 100, time aligned by means of a circular cross-correlation, and downsampled by 100.

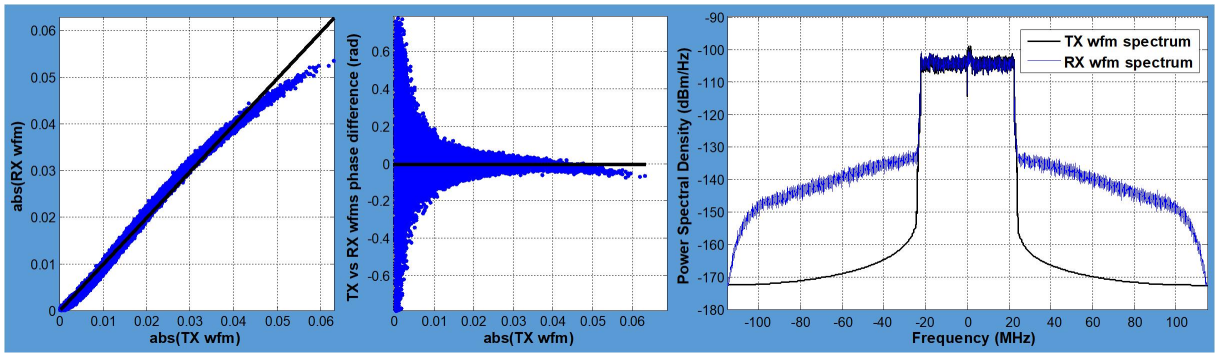


Figure 3.3: Evaluation of GaN PA linear performance (Tx versus Rx waveforms) with statistically representative test signal: AM-AM (left) and AM-PM (center) conversion and Power Spectral Density (right).

The MP, DDR-Volterra and GMP behavioral models were evaluated. GMP was chosen since it outperformed MP and DDR-Volterra in this scenario. For parameter estimation our DPD scheme followed the direct learning (DL) approach [Bra11], using the linear least squares (LS) solution as estimation method since the DPD function is non-linear but it is nevertheless linear in the parameters. DL was chosen instead of the simpler and widely used indirect learning (IL) architecture due to being better performing [Bra15], [Paa08]. The initial MATLAB Testbench and DPD architecture to combat the GaN PA non-linearities is shown in Fig. 3.4, where it can be seen that the DL parameter extraction follows a closed-loop architecture employing the feedback signal $y[n]/G_0$ (being G_0 the linear gain of the PA) together with the input signal $u[n]$ to extract the DPD coefficients. As per Section 2.2.2, the predistorted signal $x[n]$ is derived from subtracting the distortion signal $d_{GMP}[n]$ to the original signal $u[n]$, both having L samples, in the GMP behavioral modeling and DPD block.

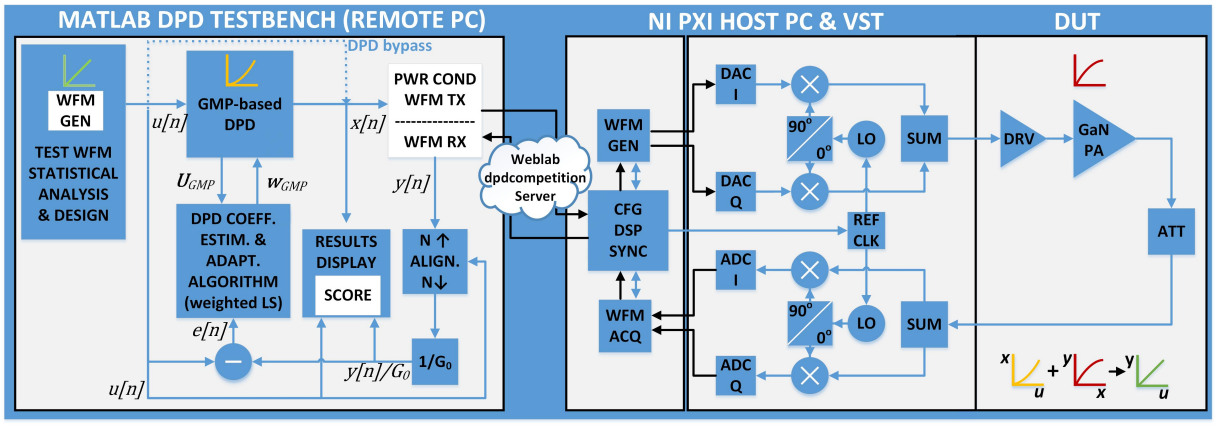


Figure 3.4: MATLAB test bench with Closed-loop adaptive DPD architecture.

With the previous architecture, multiple experiments with the Weblab test bed [Lan15] were conducted in order to tune the best GMP parameters and have the best coefficients after training the DPD for the statistically representative signals. The coefficients were stored and used against different PAPR test signals to evaluate their stability for similar PAPR range signals and the degradation suffered when signals with higher PAPR are employed. Fig. 3.5 displays the NMSE and ACPR values after conducting such measurement campaign.

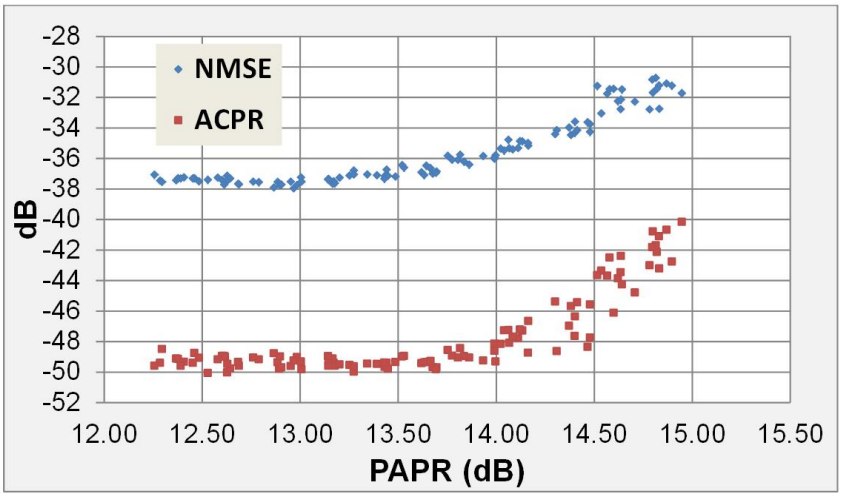


Figure 3.5: NMSE and ACPR (right) for 100 Weblab experiments with baseline DPD test bench.

Some further work was required to improve the NMSE by a few dB and push the ACPR to values below -50 dBc to increase the performance and reduce its variance. For waveforms with PAPR 13.5-14 dB, a slight degradation and increase in the variance of the performance was noticed. Beyond 14 dB PAPR values, the NMSE and more importantly the ACPR decreased significantly, all of which required the mitigation countermeasures proposed in this use case.

Techniques Targeting Performance and Results Stability

- Decomposed piecewise GMP/MP DPD modeling

The AM-AM characteristic of the PA in Fig. 3.3 shows a different behavior according to the amplitude (or alternatively, power) level. At low amplitude levels, some gain expansion can be observed, while at high amplitude levels, gain compression occurs. Since high amplitude values occur less frequently, a simplified (in terms of memory taps) behavioral model can be considered, which favors the robustness of the extracted DPD coefficients. The input signal was first decomposed into two sub-signals, as shown in Fig. 3.6 (left). Following the same principle presented in [Zhu08], we introduced piecewise curve fitting, dividing the DPD function into two segments for high and low input amplitude levels. The segment of low amplitude values was modeled with the previously detailed GMP model, while for the high amplitude values we relaxed the complexity by using a MP model. Both contributions can be accounted (i.e., since the model is linear with respect to the parameters) as seen in (3.3).

$$\boldsymbol{d}_{PW} = \boldsymbol{U}_{GMP}^{low}\boldsymbol{w}_{GMP} + \boldsymbol{U}_{MP}^{high}\boldsymbol{w}_{MP} \tag{3.3}$$

- Long-term memory effects modeling

By considering the bursty nature of the test signal, we considered the possibility of also modeling long-term memory effects. The memory effects make the output of a power amplifier at a certain time not only depend on the on the present input value, but also previous output and input values. Such bursty nature of the PA input signal may induce dynamic effects such as bias circuit modulation, self-heating and charge trapping phenomena that feature time constants much larger than the period of the RF carrier frequency, and that fall closer to the timescale of the envelope or modulation signal [Woo14], that are not tracked by the DPD coefficients unless a long term dependence function is introduced in the estimation process. Following a similar principle as that presented in [Teh12], an estimate of the long-term variable was created ($s[n]$) consisting of the average input power over a finite window with S samples. The number of samples chosen to define the window determines the long-term memory depth. The resulting vector (s) will also have L sample length.

$$s[n] = \frac{1}{S}\sum_{i=0}^{S-1}|u[n-i]|^2 \tag{3.4}$$

Considering the previous decomposed piecewise modeling and the long-term memory effects the non-linearities are modeled as follows:

$$\begin{aligned} \boldsymbol{d}_{PW,LT} &= \boldsymbol{U}_{GMP}^{low}\boldsymbol{w}_{GMP} + \boldsymbol{U}_{MP}^{high}\boldsymbol{w}_{MP} + (\boldsymbol{s}^T \cdot \boldsymbol{U}_{GMP})\boldsymbol{w}_{GMP,LT} \\ &= \boldsymbol{U}_{GMP}^{low}\boldsymbol{w}_{GMP} + \boldsymbol{U}_{MP}^{high}\boldsymbol{w}_{MP} + \boldsymbol{U}_{GMP}^{long}\boldsymbol{w}_{GMP,LT} \end{aligned} \tag{3.5}$$

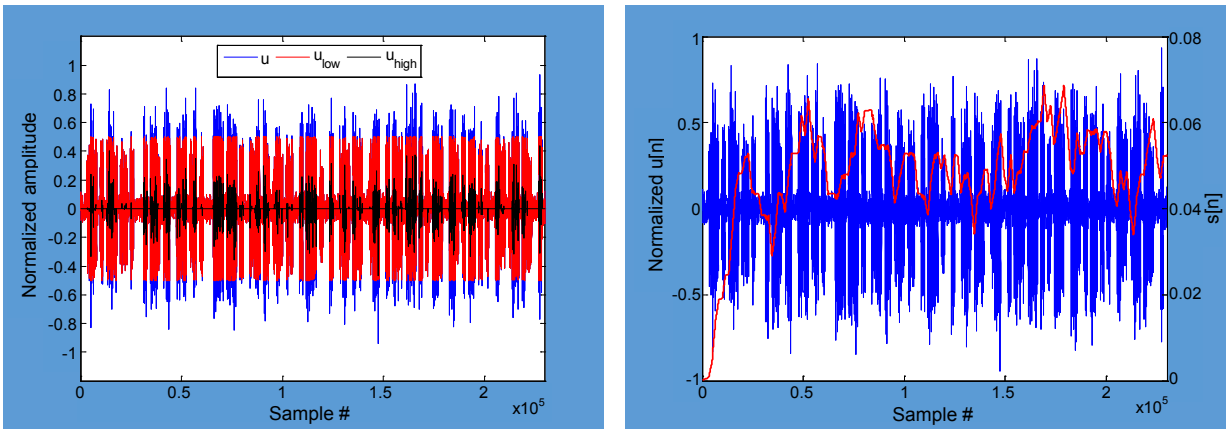


Figure 3.6: Waveform decomposition into high and low values for 0.5 amplitude threshold (left) and extraction of the long term variable (right) following the waveform power dynamics.

The number of coefficients required to obtain $\boldsymbol{d}_{PW,LT}$ and generate the predistorted signal as $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{d}_{PW,LT}$ may grow significantly when taking into account all the features described (e.g., requiring hundreds of coefficients). Model order reduction techniques like those deployed later in this chapter may be very useful to simplify the computational complexity of the DPD.

- Waveform expansion and mesh selecting

Assuming that our data Matrix has no problems of rank deficiency, we can address the accuracy problem by extending the number of observations (i.e., extending the number of required waveforms to K, see Fig. 3.8) before extracting the DPD coefficients with the LS algorithm, and then by selecting those samples that are more statistically representative of the input domain for each of the K observation waveforms.

With the histogram-based mesh-selecting (MeS) SSM in [Wan15], given a fixed length of data (rows of the data matrix) we can extend the number of observations during training, store them, and after proper selection of the most statistically representative data according to the PA AM-AM distribution, build the data Matrix providing robustness to the DPD parameter extraction. The rationale behind the MeS method consists in building a two-dimensional histogram of the PA AM-AM characteristic in order to obtain the number of samples per bin according to the probability of each segment of the mesh. Such procedure is shown in Fig. 3.7, only for illustrative purposes (the plots do not reflect its application to the use case), for a single observation ET-DPD example with 25600 training data samples and 10% selection factor (i.e., $RF_l = 10$).

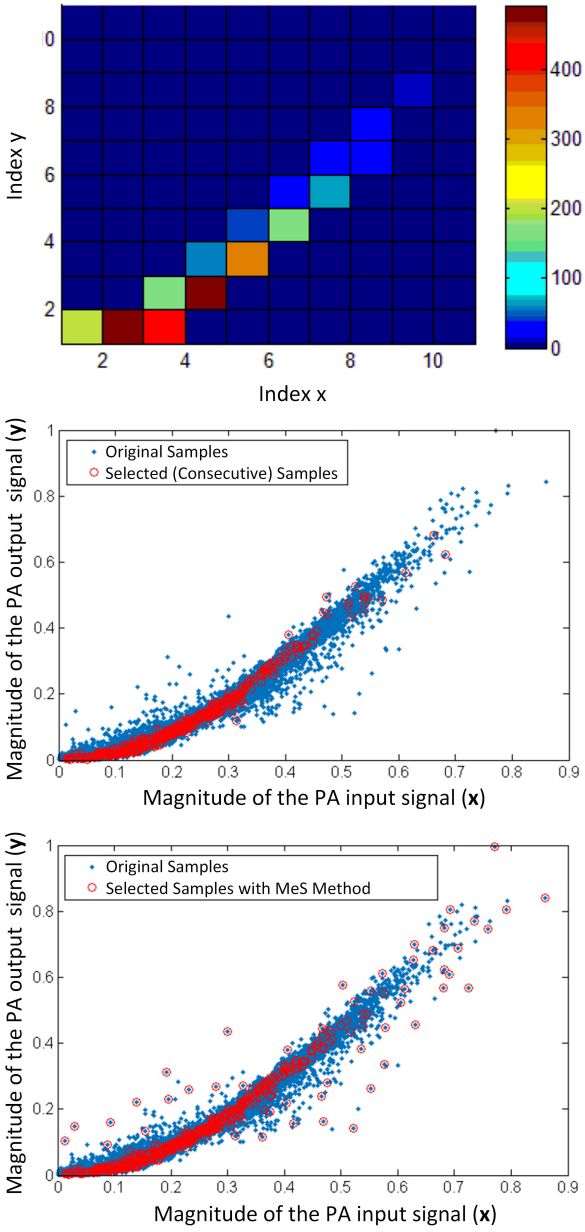


Figure 3.7: MeS ET-PA AM-AM characteristic 2D histogram with captured samples (top), AM-AM plots with initial consecutive sample selection (middle) and MeS sample selection (bottom), at 10% selection factor [Wan15].

Despite the waveform expansion (i.e., after capturing typically 3-4 waveforms per DPD iteration), since the mesh selecting sample selection technique applies a reduction factor typically higher than the number of waveforms captured, the behavioral modeling matrix will have much less rows. This will reduce the number of calculations at DPD identification and provide robust identification of coefficients since the 2D-histogram samples selected will be statistically relevant and capture with higher probability those points being affected by memory effects or strong nonlinearities, as depicted in the AM-AM plots in Fig. 3.7.

- Peak cancellation crest factor reduction

CFR was used as protection to minimize the effects of having a signal with much larger PAPR than the statistical mean value at which the predistorter has been previously trained during final tests. The peak cancellation (PC) technique shown in Section 2.3 was chosen as a clipping and filtering CFR method. The waveform statistical analysis and the NMSE versus CFR reduction for different configuration parameters is further detailed in [LB16] for a five-stage PC-CFR.

Advanced digital linearization architecture

The final DPD architecture to face the strong non-linear behavior and memory effects of a GaN PA under high PAPR OFDM-like bursty waveforms is shown in Fig. 3.8:

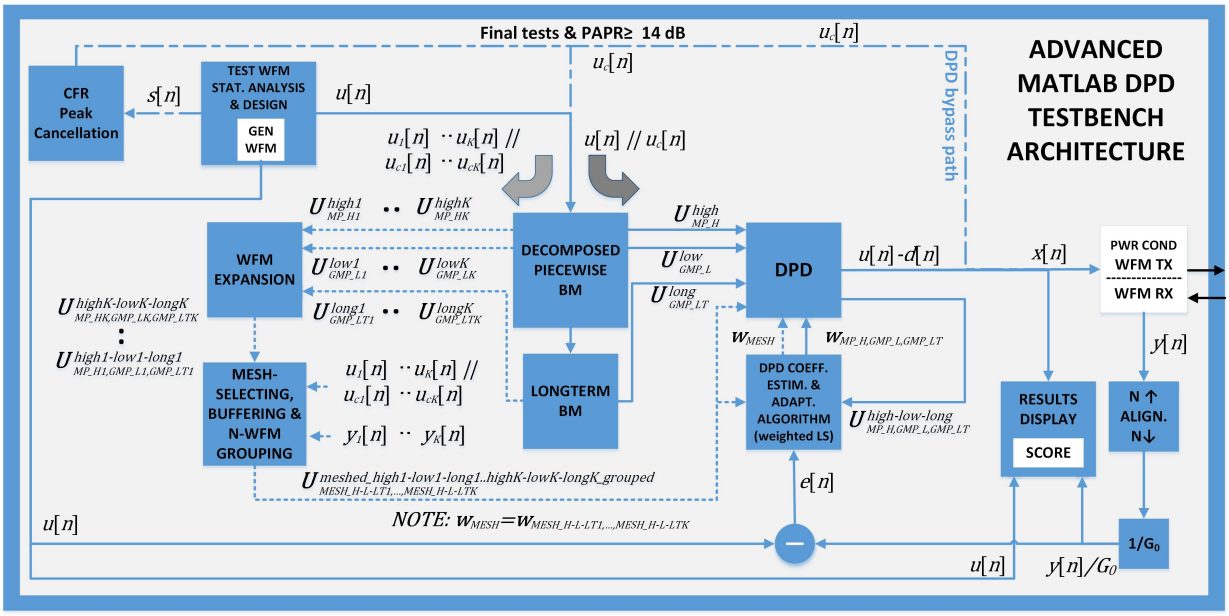


Figure 3.8: Proposed advanced DPD MATLAB test bench architecture.

With this architecture, comparing Fig. 3.5 with the new measurements in Fig. 3.9-top (both showing the results taken with 100 statistically representative test waveforms), one can see that the results show now not only much better stability (the 6 dB variation in NMSE and 9 dB variation in ACPR has been reduced to 3 dB and 2 dB, respectively), but improved performance (3.5 dB in NMSE and 1 dB in ACPR for the most frequent PAPR values).

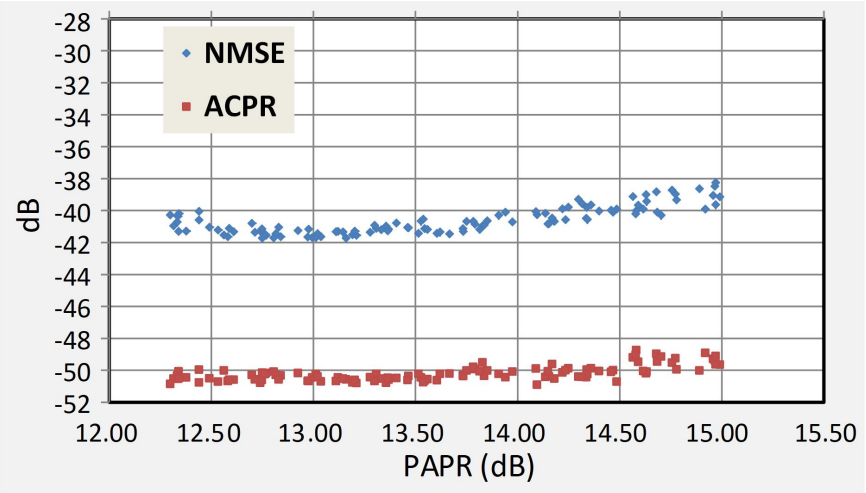


Figure 3.9: NMSE and ACPR for 100 Weblab experiments after applying advanced architecture.

3.1.3 Feature Selection Techniques for Basis Reduction in the Forward Path

Greedy Feature Selection Algorithm: Orthogonal Matching Pursuit (OMP)

When considering wide bandwidth signals, carrier aggregation or multi-band configurations [Jar18b] in high efficient transmitter architectures, such as Doherty PAs, envelope tracking PAs or outphasing transmitters, the number of parameters required in the DPD model to compensate for both static nonlinearities and dynamic memory effects can be unacceptably high. This has a negative impact in the DPD model extraction/adaptation process, because it increases the computational complexity which may provoke over-fitting and uncertainty in the DPD estimation stages [CC17]. However, by applying regularization techniques [Bra17] we can both avoid the numerical ill-conditioning of the estimation and reduce the number of coefficients of the DPD function in the forward path, which ultimately impacts the baseband processing computational complexity and power consumption. This process may be seen as a selection of the most relevant variables from a random set of original variables to provide the best fit (as illustrated in Fig. 3.10) and both make the DPD coefficients estimation less sensitive to missing data or to modeling the measurement noise.

The objective of DPD regularization is to enforce the sparsity constraint on the vector of parameters by minimizing the number of active components (i.e., ℓ_0-norm) subject to a constraint on the ℓ_2-norm squared of the identification error. Unfortunately, this is a non-deterministic polynomial-time hard (NP-hard) combinatorial search problem. Therefore, in the field of DPD linearization, several sub-optimal approaches have been proposed targeting both robust identification and model order reduction such as: the least absolute shrinkage and selection operator (LASSO), used for example by Wisell et al. in [Wis08] and consisting in a ℓ_1-norm regularization; the Ridge regression, used for example by Guan et al. in [Gua12] and consisting in a ℓ_2-norm

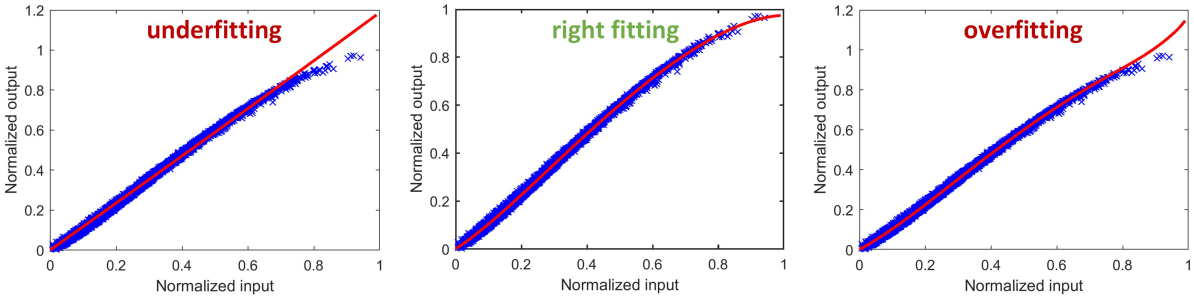


Figure 3.10: Fitting performance in the least square identification of a power amplifier model [Gil20b].

regularization; the sparse Bayesian learning (SBL) algorithm, used by Peng et al. in [Pen16a]; the orthogonal matching pursuit (OMP), a greedy algorithm for sparse approximation used in [RT15] by Reina et al. to select the most relevant basis functions of the DPD function.; an enhanced version called doubly OMP (DOMP) and a sparse model selection system based on subspace pursuit, both by Becerra et al. [Bec18b,Bec18a]; or other approaches based on decision trees such as Random Forest [Alv20]. An interesting comparison of the performance of these feature selection methods (and others classified as heuristic local search, dynamic model sizing, and global probabilistic optimization algorithms) is provided in [Bar21], where the greedy pursuits have shown to provide the best trade-off between DPD training time and linearization performance.

The OMP algorithm application procedure is shown in the digital linearizer found at Fig. 3.11. This machine learning algorithm can be used to conduct an *a priori* off-line study to select the best basis functions contributing to better modeling the PA nonlinearities for a given set of operating conditions (i.e., PA biasing point, input waveform PAPR and mean power, bandwidth related to memory effects, or temperature range). Such study, consequently, may be performed once to select the best basis in the forward path behavioral model with the goal to reduce the number of DPD coefficients and improve the conditioning and robustness of the adaptive DPD system. Such selection will remain stable if the previously mentioned operating conditions do not vary significantly.

The whole process is also described in the following:

1. Collect the input-output data of the PA with a test signal having the same statistical properties as the one that will be later used for the transmission.

2. Run off-line (e.g., in a PC running MATLAB) the OMP algorithm to obtain a sorted set of the most relevant basis functions that better characterize the PA nonlinear distortion with special emphasis on the out-of-band distortion. This search is done only once.

3. Build the closed-loop DPD taking into account the basis functions previously selected by

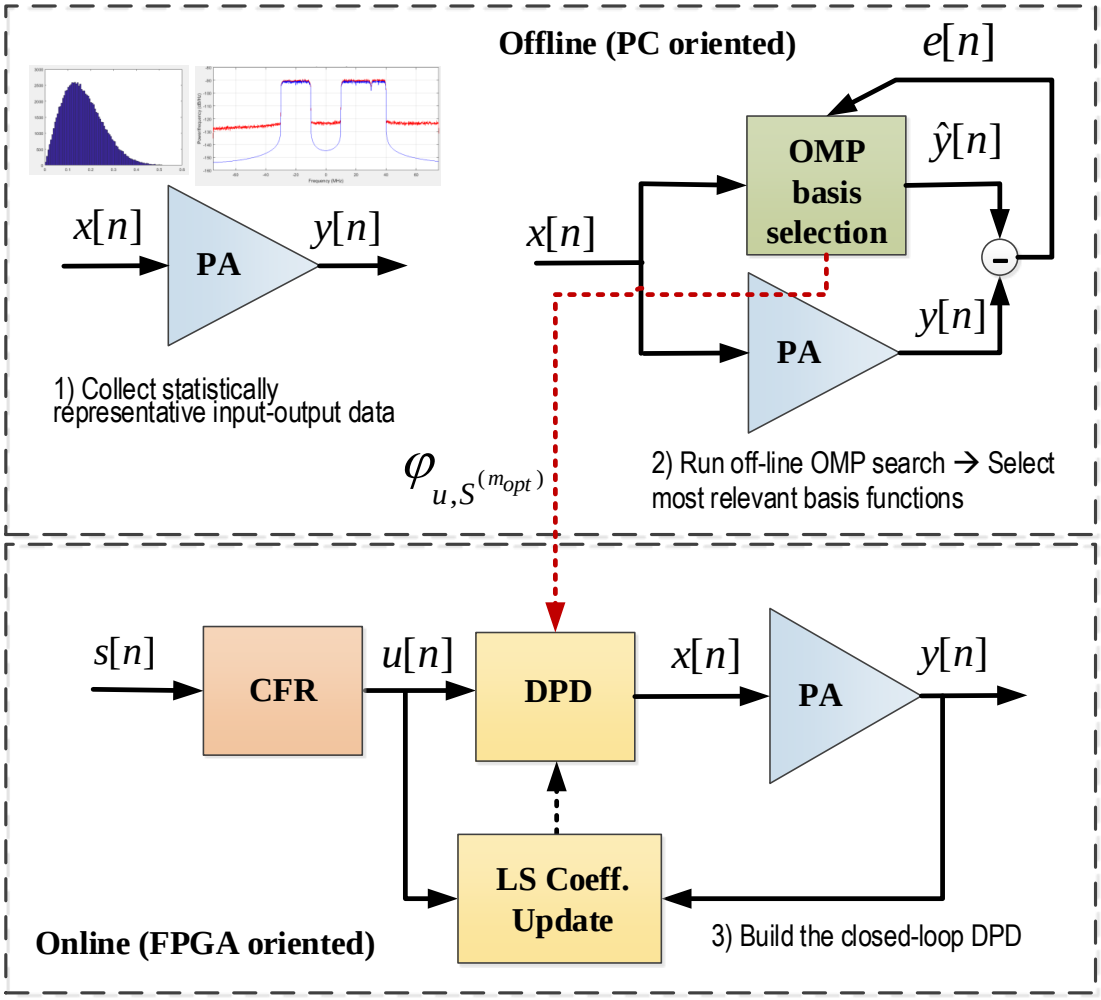


Figure 3.11: Block diagram of the OMP off-line search process and DPD linearization architecture.

the OMP algorithm. The identification of the DPD will be well-conditioned, robust and may take only few iterations to converge.

A more specific definition of the OMP algorithm follows. Following the notation in Fig. 3.11 and accounting also for that introduced in Section 2.2, the estimated PA behavioral model output $\hat{\mathbf{y}}$ can be defined as:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{w} \tag{3.6}$$

where $\boldsymbol{w}$ is the $N \times 1$ vector with the estimated coefficients for PA modeling and $\boldsymbol{X}$ is the $L \times N$ data matrix containing the basis functions or components.

As explained before, the objective is to minimize the number of active components (i.e.,

ℓ_0-norm) subject to a constraint on the ℓ_2-norm squared of the identification error,

$$\begin{aligned} &\min_{\boldsymbol{w}} \|\boldsymbol{w}\|_0 \\ \text{subject to} \quad &\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|_2^2 \leq \varepsilon \end{aligned} \tag{3.7}$$

where $\boldsymbol{y}$ is the $L \times 1$ vector of measured data at the PA output.

In order to minimize the number of coefficients being required by the DPD function in the forward path, we can assume that the optimal subset of selected basis functions of the DPD function will be the same as that used for modeling the PA behavior. To overcome the NP-hard combinatorial search problem of such coefficient minimization, OMP is operated as shown in Algorithm 4:

Algorithm 4 Orthogonal Matching Pursuit

procedure OMP ($\boldsymbol{y}$,$\boldsymbol{X}$,RF_d)
  initialization:
    $\boldsymbol{e}^{(0)} = \boldsymbol{y} - \hat{\boldsymbol{y}}^{(0)}$; with $\hat{\boldsymbol{y}}^{(0)} = 0$
    $\boldsymbol{\Upsilon}^{(0)} \leftarrow \{\}$
    $\boldsymbol{X}^{(0)} \leftarrow \boldsymbol{X}_{L\times N}$
  for $m = 1$ **to** N/RF_d **do**
    $\boldsymbol{X}_{\{i\}}^{(m-1)} \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\leftarrow} \frac{\boldsymbol{X}_{\{i\}}^{(m-1)}}{\|\boldsymbol{X}_{\{i\}}^{(m-1)}\|_2}$
    $i^{(m)} \leftarrow \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmin}} \min_{\boldsymbol{w_i}} \|\boldsymbol{e}^{(m-1)} - \boldsymbol{X}_{\{i\}}^{(m-1)}\boldsymbol{w_i}\|_2^2 \approx \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmax}} |{\boldsymbol{X}_{\{i\}}^{(m-1)}}^H \boldsymbol{e}^{(m-1)}|$
    $\boldsymbol{\Upsilon}^{(m)} \leftarrow \boldsymbol{\Upsilon}^{(m-1)} \cup i^{(m)}$
    $\boldsymbol{w} \leftarrow \left(\boldsymbol{X}_{\{\boldsymbol{\Upsilon}^{(m)}\}}^H \boldsymbol{X}_{\{\boldsymbol{\Upsilon}^{(m)}\}}\right)^{-1} \boldsymbol{X}_{\boldsymbol{\Upsilon}^{(m)}\}}^H \boldsymbol{y}$
    $\hat{\boldsymbol{y}}^{(m)} \leftarrow \boldsymbol{X}_{\{\boldsymbol{\Upsilon}^{(m)}\}} \boldsymbol{w}$
    $\boldsymbol{e}^{(m)} = \boldsymbol{y} - \hat{\boldsymbol{y}}^{(m)}$
  end for
  return $\boldsymbol{\Upsilon}$
  return $\boldsymbol{X}_{\{\boldsymbol{\Upsilon}\}}$
end procedure

In Algorithm 4, $\boldsymbol{\Upsilon}^{(m)}$ is the support set that contains the indices of the basis functions that better describe the PA behavioral model. Being N the total number of basis and RF_d the dimensionality reduction factor (the d subscript refers to dimensionality reduction) applied to the basis selection, $\boldsymbol{\Upsilon}$ will finally contain N/RF_d indices that will be later applied to matrix $\boldsymbol{U}$ when calculating the DPD coefficients. The basis functions better contributing to the residual error minimization (between the measured output signal and its modeled version) are selected and added to $\boldsymbol{\Upsilon}^{(m)}$ at every iteration of the OMP search. In order to simplify the index $i^{(m)}$ calculation, the elements of $\boldsymbol{X}_{\{\boldsymbol{\Upsilon}^{(m)}\}}$ are normalized in power. By maximizing the absolute value of the correlation between the basis function $\boldsymbol{X}_{\{i\}}$ and the residual error $\boldsymbol{e}^{(m-1)}$ of the previous iteration, such index calculation is attained. In $\boldsymbol{\Upsilon}^{(m)}$ these indices referring to basis functions or

active components are sorted according to their contribution or relevance. If instead of defining a basis reduction factor and a search for a finite number of iterations (N/RF_d), the optimum number of basis or coefficients is wanted (i.e., being smaller than N), then some information criterion, such as the Akaike (AIC) or the Bayesian (BIC) [RT15] can be used to stop the index search in Algorithm 4 at $m_{opt} < N$. Finally, the subset of selected basis functions determined by $\boldsymbol{\Upsilon}$ and applied to $\boldsymbol{X}_{\{\boldsymbol{\Upsilon}^{(m)}\}}$ to either reduce complexity or improve conditioning, will then be used as $\boldsymbol{U}_{\{\boldsymbol{\Upsilon}^{(m)}\}}$ (either for $m = N/RF_d$ or $m = m_{opt}$) to carry out the DPD processing as per shown in Section 2.2.2 (i.e., assuming now that $\boldsymbol{U}^{OMP} = \left(\boldsymbol{\phi_u}^{OMP}[0], \boldsymbol{\phi_u}^{OMP}[1], \cdots, \boldsymbol{\phi_u}^{OMP}[L-1]\right)^T$, where $\boldsymbol{\phi_u}^{OMP} = \boldsymbol{\phi}_{\boldsymbol{u}\{\boldsymbol{\Upsilon}^{(m)}\}}$).

Finally, it is noted that in the DPD domain just a few works have been proposed as enhancements or adaptations made over the OMP algorithm. One example outperforming OMP both in in-band and out-of-band modeling and DPD performance is the previously introduced DOMP algorithm. This algorithm is detailed in Section 4.2.2, and is applied for the first time to reduce the input dataset features of a SISO neural network predistorter. Another example focusing only on providing enhanced out-of-band modeling performance is the spectral weighting OMP (SW-OMP) [Gil18] which is further detailed in the use case that follows.

Feature Selection Use Case

Spectral Weighting OMP algorithm for trading off EVM vs ACPR in DPD

This use case is aimed at providing an alternative approach to the classical OMP algorithm which is based on a spectral weighting strategy. This new approach is oriented to select the most relevant DPD basis functions by paying special attention to the out-of-band distortion compensation. When the mean squared error is mainly dominated by the linear part, by adding spectral weighting (through in-band notching of the reference signal and the DPD basis functions) the OMP algorithm can focus on the subset of parameters that better contribute to minimize the mean squared error which is dominated by the out-of-band distortion.

The objective is to build a robust and accurate DPD linearizer that meets the linearity levels specified in communications standards making use of the minimum necessary number of coefficients. Reducing the number of coefficients has a beneficial impact on the amount of FPGA resources [Woo17] (e.g., less DSP48 slices and RAM memory blocks in Xilinx FPGAs) required for the implementation of the DPD in the forward path. An off-line study of the most relevant basis functions involved in the characterization of the PA nonlinear distortion is conducted in advance in order to be later used for DPD purposes as previously described. The search of the most relevant components, given a general behavioral model descriptor (e.g., full Volterra series, generalized memory polynomial, etc.), is carried out only once in an off-line process. To further enhance the DPD linearization performance obtained with the subset of the selected components, the SW-OMP algorithm is oriented toward emphasizing the out-of-band distortion

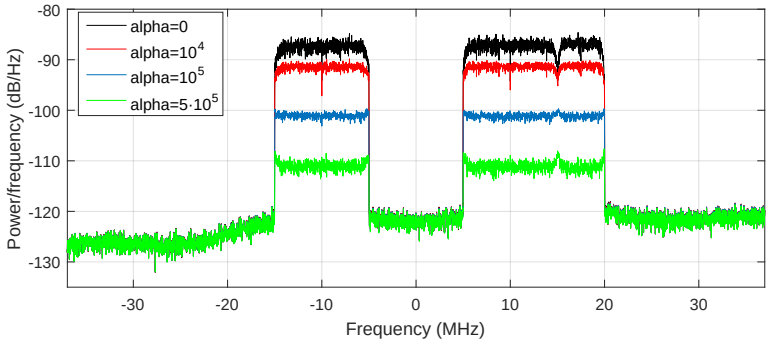


Figure 3.12: Spectra of the filtered PA outputs for different alpha values.

part of the mean squared error.

Description of the SW-OMP algorithm

The PA behavioral modeling identification is conducted as per described in Section 2.2.1 for the GMP model. The SW-OMP algorithm capable is described in order to find the most relevant basis functions from matrix $\boldsymbol{X} = \left(\boldsymbol{\phi_x}[0], \boldsymbol{\phi_x}[1], \cdots, \boldsymbol{\phi_x}[L-1]\right)^T$ where $\boldsymbol{\phi_x}[n] = \left(\varphi_1^x[n], \cdots, \varphi_i^x[n], \cdots, \varphi_O^x[n]\right)^T$ is now the $O \times 1$ vector of basis functions $\varphi_i^x[n]$ (with $i = 1, \cdots O$) at time n. This basis selection will be then applied to matrix $\boldsymbol{U}$ in the DPD linearizer.

In order to select the most relevant basis functions by paying special attention to the out-of-band distortion compensation, the proposed SW-OMP targets the minimization of an in-band notched version of the mean square error. Consequently, the PA output and all the basis functions will be notched in-band. Therefore, as described in [Bra12], the PA output signal is filtered in the frequency domain as follows,

$$Y_f[k] = \Gamma\{Y[k]\} = \frac{Y[k]}{1 + \alpha|Y[k]|^2} \tag{3.8}$$

with α being a parameter to control the notching level (see Fig. 3.12) and with $Y[k]$ being the discrete Fourier transform (DFT) of $y[n]$,

$$Y[k] = DFT_N\{y[n]\} = \sum_{n=0}^{N-1} y[n] e^{-j\frac{2\pi}{N}kn} \tag{3.9}$$

where $k = 0, \cdots N-1$ with $N \geq L$ (note that the total number of basis goes now from 1 to O to avoid confusion with the classical DFT formulation). Similarly, each basis function in the frequency domain is filtered as follows,

$$\boldsymbol{Z}_{\boldsymbol{x},f}[k] = \Gamma\{\boldsymbol{Z_x}[k]\} = \left(\Gamma\{\zeta_1^x[k]\}, \cdots, \Gamma\{\zeta_i^x[k]\}, \cdots, \Gamma\{\zeta_O^x[k]\}\right)^T$$

with $\zeta_i^x[k]$ $(i = 1, \cdots, O)$ being the DFT of $\varphi_i^x[n]$ and $\Gamma\{\cdot\}$ the filtering operator. In the frequency domain, the $N \times O$ matrix of filtered basis functions is defined as

$$\boldsymbol{\Psi}_{\boldsymbol{x},f} = \Big(\boldsymbol{Z}_{x,f}[0], \cdots, \boldsymbol{Z}_{x,f}[k], \cdots, \boldsymbol{Z}_{x,f}[N-1]\Big)^T. \tag{3.10}$$

Now, $\boldsymbol{\Upsilon}^{(m)}$ is again defined as the support set containing the indices of the basis functions building the model. Considering that m_{max} is the number of basis functions under study (i.e., $m_{max} = O$), the proposed SW-OMP algorithm is defined in Algorithm 5.

At every iteration, the basis function that better contributes to minimize the filtered residual error is selected and added to the support set $\boldsymbol{\Upsilon}^{(m)}$. Again, the elements of $\boldsymbol{\Psi}_{\boldsymbol{x},f\{\boldsymbol{\Upsilon}^{(m)}\}}$ have been normalized in power to simplify the index $i^{(m)}$ calculation in line 6 of the algorithm, which can be obtained by maximizing the absolute value of the correlation between the filtered basis function $\boldsymbol{\Psi}_{\boldsymbol{x},f,\{i\}}$ $\big($i.e, $\boldsymbol{\Psi}_{\boldsymbol{x},f,\{i\}} = \Gamma\big\{DFT_N\{\boldsymbol{X}_{\{i\}}\}\big\}\big)$ and the filtered residual error $\mathbf{E}^{(m-1)}$ of the previous iteration. After a complete SW-OMP search, we obtain a vector $\boldsymbol{\Upsilon}^{(m)}$ with the indices of all the original basis functions (active components) sorted according to their relevance. Again, by using some information criterion, such as AIC or BIC, it is possible to determine the optimum number of coefficients (m_{opt}), where $m_{opt} < m_{max}$. Finally, the subset of selected basis functions for model identification $\boldsymbol{X}_{\{\boldsymbol{\Upsilon}^{(m_{opt})}\}}$, will be also used as $\boldsymbol{U}_{\{\boldsymbol{\Upsilon}^{(m_{opt})}\}}$ used to carry out the DPD as previously explained for the OMP algorithm.

Algorithm 5 Spectral Weighting Orthogonal Matching Pursuit

1: **procedure** SW-OMP ($\boldsymbol{Y}_f$,$\boldsymbol{\Psi}_{\boldsymbol{x},f}$,$m_{max}$)
2: initialization:
3: $\boldsymbol{E}_f^{(0)} = \boldsymbol{Y}_f - \hat{\boldsymbol{Y}}_f^{(0)}$; with $\hat{\boldsymbol{Y}}_f^{(0)} = 0$
4: $\boldsymbol{\Upsilon}^{(0)} \leftarrow \{\}$
5: $\boldsymbol{\Psi}_{\boldsymbol{x},f}^{(0)} \leftarrow \boldsymbol{\Psi}_{\boldsymbol{x},f_{N\times O}}$
6: **for** $m = 1$ **to** m_{max} **do**
7: $\boldsymbol{\Psi}_{\boldsymbol{x},f\{i\}}^{(m-1)} \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\longleftarrow} \frac{\boldsymbol{\Psi}_{\boldsymbol{x},f\{i\}}^{(m-1)}}{\|\boldsymbol{\Psi}_{\boldsymbol{x},f\{i\}}^{(m-1)}\|_2}$
8: $i^{(m)} \leftarrow \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmin}} \min_{\boldsymbol{w_i}} \|\boldsymbol{E}_f^{(m-1)} - \boldsymbol{\Psi}_{\boldsymbol{x},f\{i\}}^{(m-1)}\boldsymbol{w_i}\|_2^2 \approx \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmax}} |\boldsymbol{\Psi}_{\boldsymbol{x},f\{i\}}^{(m-1)^H}\boldsymbol{E}_f^{(m-1)}|$
9: $\boldsymbol{\Upsilon}^{(m)} \leftarrow \boldsymbol{\Upsilon}^{(m-1)} \cup i^{(m)}$
10: $\boldsymbol{w} \leftarrow \big(\boldsymbol{\Psi}_{\boldsymbol{x},f\{\boldsymbol{\Upsilon}^{(m)}\}}^H \boldsymbol{\Psi}_{\boldsymbol{x},f\{\boldsymbol{\Upsilon}^{(m)}\}}\big)^{-1} \boldsymbol{\Psi}_{\boldsymbol{x},f\{\boldsymbol{\Upsilon}^{(m)}\}}^H \boldsymbol{Y}_f$
11: $\hat{\boldsymbol{Y}}_f^{(m)} \leftarrow \boldsymbol{\Psi}_{\boldsymbol{x},f\{\boldsymbol{\Upsilon}^{(m)}\}}\boldsymbol{w}$
12: $\boldsymbol{E}_f^{(m)} = \boldsymbol{Y}_f - \hat{\boldsymbol{Y}}_f^{(m)}$;
13: **end for**
14: **return** $\boldsymbol{\Upsilon}$
15: **return** $\boldsymbol{\Psi}_{\boldsymbol{x},f\{\boldsymbol{\Upsilon}\}}$
16: **end procedure**

The benefit of using the SW-OMP is to reduce the computational complexity of the DPD behavioral model in the forward path. The SW-OMP algorithm is inherently stable and the

index search is repeated until the upper-bound m_{max} on the number of active coefficients is reached, or until the norm of the filtered residual error $\left\|\mathbf{E}^{(m)}\right\|_2$ falls below the noise tolerance threshold. The running time depends on both the number of coefficients considered in the search and on the digital signal processor used for running the algorithm. However, by comparing the SW-OMP algorithm with the Less Relevant Basis Removal (LRBR) brute-force technique presented in [Gil16] in terms of computational time, the OMP is 21 times faster, while the accuracy of the search is similar to that obtained with OMP. The computational complexity of the SW-OMP in comparison to the OMP is slightly higher due to the FFT transformations and filtering operations. However, when running an off-line SW-OMP search (e.g., in a PC running MATLAB instead of in a FPGA) considering a high number of coefficients (e.g., > 300 coeff.), the computational time devoted to calculating the Fourier transforms and filtering operations is negligible.

Instead, thanks to the SW-OMP search, some benefits in the DPD forward path are obtained in terms of robustness (e.g., fast convergence in just 3 iterations when considering the use of different input data featuring different PAPR), flexibility (e.g., trade-off between the number of selected coefficients and the DPD linearization performance) and reliability (by selecting the most relevant basis the over-parametrization is avoided which yields to a properly conditioned and accurate coefficient estimation).

Experimental Results

The test bench shown in Fig. 3.13 was used to conduct experimental validation of the SW-OMP technique. To emulate a LTE carrier-aggregation transmission, two 20 MHz bandwidth channels and one 10 MHz channel were grouped as shown in Fig. 3.12. The PAPR of the waveform has been limited to 9.5 dB by using the peak cancellation crest factor reduction technique. The DPD processing blocks run in a PC with MATLAB, which is also employed to interface the waveform generation and acquisition instruments. The direct RF generation of the LTE carrier-aggregated signal was carried out through the arbitrary waveform generator M8190A from Keysight, considering a clock rate of 7.968 GHz and 14 bits. The device under test was a class AB-biased GaN pHEMT 6 W PA delivering 28.6 dBm mean output power at 2140 MHz (downlink base-station LTE band). The gate bias was chosen to maximize efficiency at the price of showing severe nonlinear behavior. Finally, a digital storage oscilloscope (DSO Keysight 90404A) was used to acquire the RF output signal with 8-bit resolution and at sampling rate of 20 GS/s.

For the experimental results, we considered a GMP behavioral model with a configuration of polynomial orders and memory terms that resulted in an initial dictionary of $m_{max} = 596$ basis functions. The use of 596 coefficients in the GMP model resulted in an ill-conditioned estimation. By using the Moore-Penrose inverse (i.e., $(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{U}^H$) the coefficients estimation

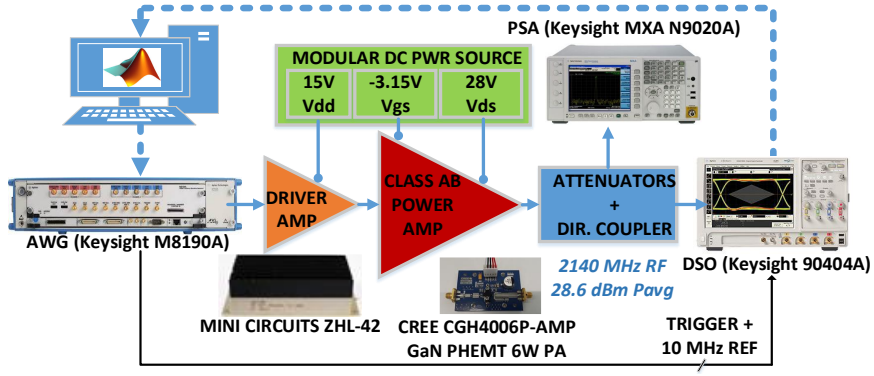


Figure 3.13: Experimental test bench employed for SW-OMP DPD profiling.

was inaccurate and some regularization was required. With the SW-OMP we addressed the over-parametrization problem by properly selecting the most relevant basis functions that minimize the out-of-band distortion.

After applying the SW-OMP search and the BIC approximation, we obtained a reduced set (being sorted according to their relevance) of around $m_{opt} = 350$ active components, which resulted in a perfectly well-conditioned estimation. In addition, as shown in Fig. 3.14, further pruning was considered by selecting some of the most relevant components of $\boldsymbol{\Upsilon}^{(m_{opt})}$ and by considering different notching levels (i.e., different α values in (3.8) as shown in Fig. 3.12). We empirically found that $\alpha = 5 \cdot 10^5$ is the maximum notching level that guarantees the minimum in-band level required for accurate selection of the most relevant basis. The DPD linearization performance was evaluated in terms of ACPR and NMSE. As expected (see Fig. 3.14), under severe nonlinear behavior of the PA, the selection made by using the SW-OMP with the highest α outperforms the classical OMP approach and is more robust against the inherent ACPR degradation suffered when reducing the number of coefficients of the DPD function.

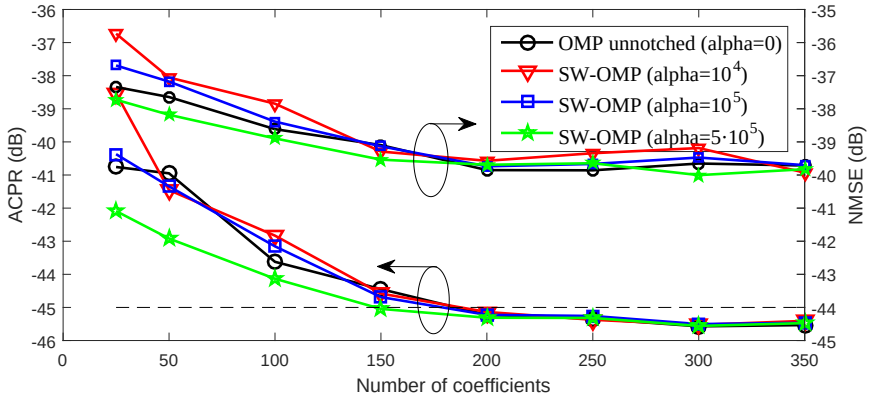


Figure 3.14: ACPR and NMSE versus number of DPD selected coefficients (using SW-OMP) for different alpha values of the notch filter.

In addition, to show the convergence speed and robustness of the closed-loop DPD when

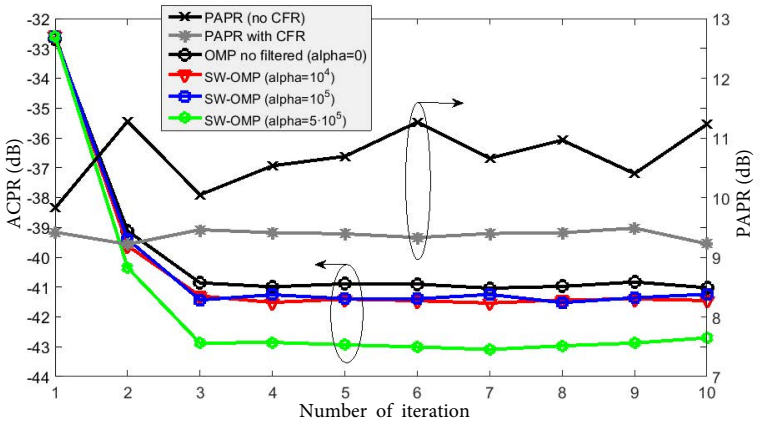


Figure 3.15: PAPR and ACPR vs # of iterations, with 50 coeff. SW-OMP DPD.

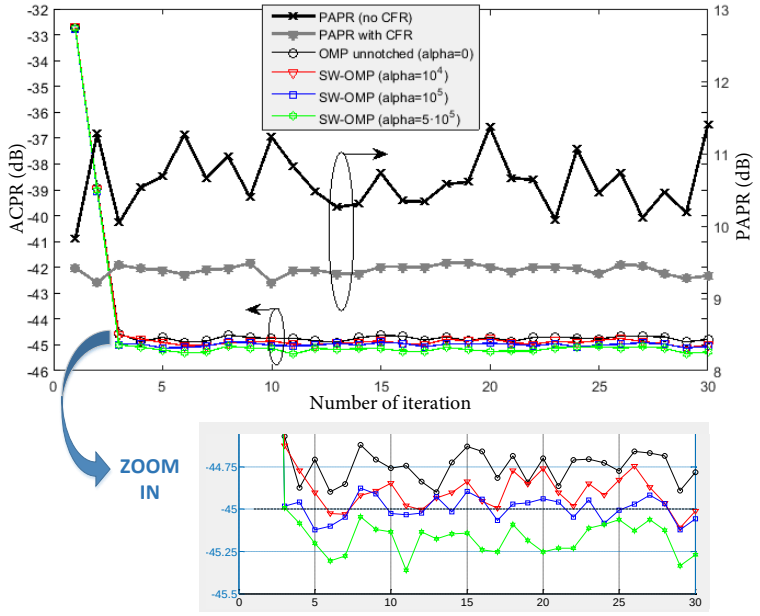


Figure 3.16: PAPR and ACPR vs # of iterations, with 150 coeff. SW-OMP DPD.

considering different data bursts with different PAPR values at each iteration, Fig. 3.15 and Fig. 3.16 show, respectively, the linearization performance of the DPD operating with 50 and 150 properly selected coefficients and considering different notching levels. By taking into account this particular PA nonlinear behavior, the target of -45 dBc of ACPR is only met when using a minimum of 150 coefficients of the GMP behavioral model (see Fig. 3.16). Fig. 3.17 shows the output power spectra before and after DPD linearization when considering 150 coefficients being properly selected with the SW-OMP algorithm and when considering the highest notching level (i.e., $\alpha = 5 \cdot 10^5$). Nevertheless, the advantage given by conducting a previous SW-OMP search in comparison to the classical OMP search, is more evident when fewer coefficients are considered in the DPD (e.g., only 50 coefficients in Fig. 3.15) to compensate for the PA severe nonlinear behavior.

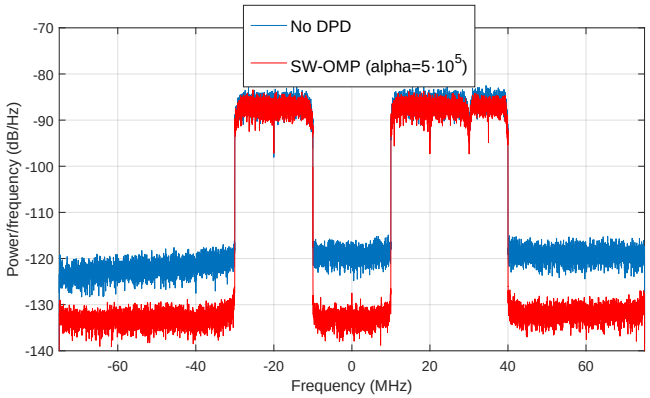


Figure 3.17: PA output power spectra before and after SW-OMP DPD.

To conclude this Feature Selection Use Case, the proposed SW-OMP algorithm emphasizes the importance of the out-of-band distortion in the OMP residual error, forcing the selection of the DPD components that better compensate for the adjacent channel spectral regrowth. Experimental results showed that, thanks to the proposed SW-OMP technique, the inherent ACPR and NMSE degradation suffered when significantly reducing the number of coefficients can be mitigated (and improved in comparison to the classical OMP under severe PA nonlinear behavior), while the robustness of the DPD is enhanced, since no further adaptation is required after 3 iterations.

3.1.4 Feature Extraction Techniques for Basis Reduction in the Feedback Path

Principal Component Analysis (PCA)

Section 3.1.3 has shown that when targeting a field-programmable gate array (FPGA) implementation, the DPD function in the forward path should be designed as simple as possible (i.e., including the minimum and most relevant basis functions) to save as many hardware logic resources and memory as possible. A lot of effort has been dedicated in the literature to propose efficient digital predistortion (DPD) architectures, either polynomial-based or LUT based , in order to implement the forward path of the digital predistorter in a real-time digital signal processor. The processing capabilities and the memory of FPGAs have increased dramatically over time. Despite this fact, having to deal with parametric models with a huge number of coefficients is not desirable, not only because of its negative impact on the FPGA resources utilization, but also because of its extraction/adaptation may derive to overfitting and uncertainty. By applying regularization, it is possible to both avoid the numerical ill-conditioning of the estimation and reduce the number of coefficients of the DPD function which ultimately impacts the baseband processing computational complexity and power consumption. Little attention, however, is paid to the identification/adaptation subsystem that most of the time is addressed in the literature by

simply solving the least squares (LS) regression applying the Moore-Penrose inverse to extract the DPD parameters as per (2.19).

In the latter specific domain, the principal component analysis (PCA) theory [S.01, Hot33] can also be applied to transform the original matrix with the DPD basis functions $\boldsymbol{U}$ in order to reduce the computational complexity at the DPD feedback or update path or prevent ill-conditioning in complex nonlinear systems where the complexity scales exponentially. PCA is a statistical learning technique used to convert an original basis of eventually correlated features or components into a new basis set of uncorrelated orthogonal components called principal components. These are linear combinations of the original variables that target capturing the maximum variance in the data. As explained in [Gil13] and [LB18], this is obtained through a change of basis using a transformation matrix, now defined as $\boldsymbol{V}$, which is originally an $N \times N$ dimensional matrix since it contains the eigenvectors of the covariance matrix of $\boldsymbol{U}$, where

$$cov(\boldsymbol{U}) = \frac{1}{L-1}\big((\boldsymbol{U} - E\{\boldsymbol{U}\})^H(\boldsymbol{U} - E\{\boldsymbol{U}\})\big) \approx \boldsymbol{U}^H\boldsymbol{U} \tag{3.11}$$

and $E\{\cdot\}$ is the expected value. The columns of the $L \times N$-dimensional matrix $\boldsymbol{U}$ contain the DPD basis functions $\boldsymbol{\varphi_i}$ with $i = 1, \cdots, N$. The principal components of the basis functions in $\boldsymbol{U}$ are defined as the eigenvectors of $\boldsymbol{U}\boldsymbol{U}^H$. Nevertheless, as proved in [Jol02], $\boldsymbol{U}^H\boldsymbol{U}$ and $\boldsymbol{U}\boldsymbol{U}^H$ have the same eigenvalues and, moreover, their eigenvectors are related as described in the following,

$$(\boldsymbol{U}^H\boldsymbol{U})\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i \rightarrow (\boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{U}\boldsymbol{v}_i = \lambda_i\boldsymbol{U}\boldsymbol{v}_i \tag{3.12}$$

with $\boldsymbol{v}_i$ being the i^{th} eigenvector of $\boldsymbol{U}^H\boldsymbol{U}$. For each i,

$$(\boldsymbol{U}^H\boldsymbol{U})\boldsymbol{V} = \boldsymbol{\lambda}\boldsymbol{V} \rightarrow (\boldsymbol{U}\boldsymbol{U}^H)\boldsymbol{U}\boldsymbol{V} = \boldsymbol{\lambda}\boldsymbol{U}\boldsymbol{V} \tag{3.13}$$

where $\boldsymbol{V} = \Big(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_i, \cdots, \boldsymbol{v}_N\Big)$ is the $N \times N$ transformation matrix. The linear combination $\boldsymbol{U}\boldsymbol{V}$, corresponds to the eigenvectors of $\boldsymbol{U}\boldsymbol{U}^H$, which are therefore the desired principal components of the DPD basis functions (i.e., columns) in $\boldsymbol{U}$. Moreover, $\boldsymbol{\lambda}$ is the diagonal matrix containing the eigenvalues of both the $\boldsymbol{U}\boldsymbol{U}^H$ and the $\boldsymbol{U}^H\boldsymbol{U}$ matrices.

$\hat{\boldsymbol{U}}$ is defined as the PCA-transformed matrix which is therefore obtained as

$$\hat{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{V}, \tag{3.14}$$

where $\hat{\boldsymbol{U}} = \Big(\boldsymbol{\Phi}[0], \cdots, \boldsymbol{\Phi}[n], \cdots, \boldsymbol{\Phi}[L-1]\Big)^T$ will be an $L \times N$ data matrix, and where $\boldsymbol{\Phi}^T[n] = \Big(\psi_1[n], \cdots, \psi_j[n], \cdots, \psi_N[n]\Big)$ is the $1 \times N$ data vector containing the new orthogonal basis functions or principal components $\psi_j[n]$ (with $j = 1, \cdots, N$) at time n.

PCA model order reduction

A graphical representation of the PCA transformation applied to DPD is shown in Fig. 3.18:

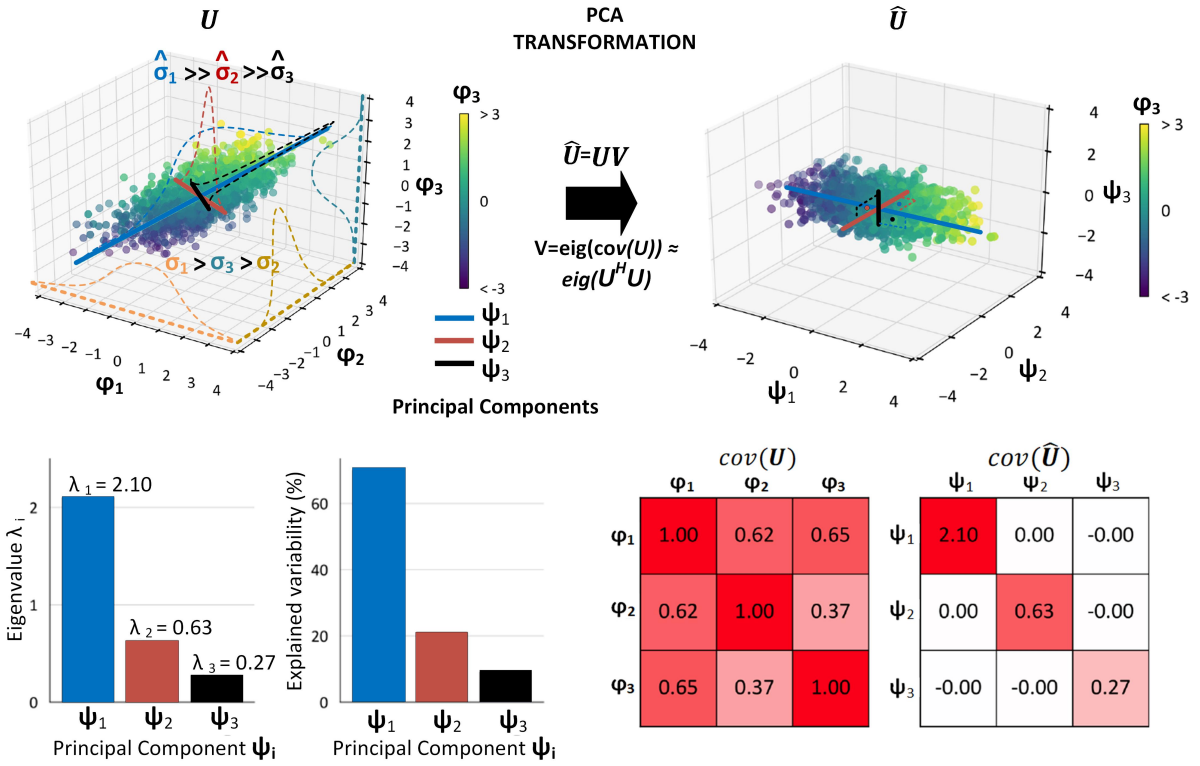


Figure 3.18: PCA transformation considering three-dimensional data (adapted from [Bar17a]).

Considering now three-dimensional data, the original coordinate axes are formed by the $\boldsymbol{\varphi_i}$ basis functions, with $i = 1, \cdots, 3$, that belong to the $\boldsymbol{U}$ behavioral modeling matrix and that will be applied to the L data samples. The new coordinate axes, belonging to the $\hat{\boldsymbol{U}}$ PCA-transformed matrix, are now $\boldsymbol{\psi_j}$ with $j = 1, \cdots, 3$, corresponding to the three eigenvectors or directions of highest variance of the original data. As it can be seen in Fig. 3.18 (upper-left corner) the first eigenvector $\boldsymbol{\psi_1}$ has a significantly larger variance than the second and third eigenvectors ($\boldsymbol{\psi_2}$ and $\boldsymbol{\psi_3}$, respectively), whereas in the original coordinate axes, the differences in the variances of the data on the three axes was not so significantly different. For geometrical illustration purposes, the transformation matrix $\boldsymbol{V}$ may be thought of as a linear rotation matrix providing new basis where the correlated data is made orthogonal and thus removable. If we now removed the $\boldsymbol{\psi_2}$ and $\boldsymbol{\psi_3}$ eigenvectors (or principal components) and project the data on the dimension of $\boldsymbol{\psi_1}$, the loss of information would be considerably shortened compared to the loss that would be obtained with the original coordinates.

Given the orthogonality of the new transformed basis functions in $\hat{\boldsymbol{U}}$ and the fact that these components sorted according to their relevance (i.e., the most relevant components showing higher eigenvalue are the ones appearing first and so on), it is possible to apply DPD model order reduction in the feedback path by just removing the components (or columns of $\hat{\boldsymbol{U}}$) which have the smallest eigenvalue. Sample applies to matrix $\boldsymbol{V}$ since the eigenvalues of $\boldsymbol{U}^H\boldsymbol{U}$ and

$\boldsymbol{U}\boldsymbol{U}^H$ are the same, and their eigenvectors are related as per (3.13).

Therefore, what makes sense is to apply the reduction to matrix $\boldsymbol{V}$ accounting for the properly sorted relevant eigenvalue contributions and now account for a reduced transformation matrix $\boldsymbol{V_{rdx}} = \left(\boldsymbol{v_1}, \cdots, \boldsymbol{v_i}, \cdots \boldsymbol{v_{N_{rdx}}}\right)$, where $N_{rdx} < N$ ($\boldsymbol{V_{rdx}}$ will be $N \times N_{rdx}$ dimensional) can define the last component or column being considered from $\boldsymbol{V}$, or be linked to a specific dimensionality reduction factor RF_d as $N_{rdx} = \frac{N}{RF_d}$. The resulting PCA-transformed matrix will be now $\boldsymbol{U^{PCA}} = \hat{\boldsymbol{U}} = \boldsymbol{U}\boldsymbol{V_{rdx}}$, and feature $L \times N_{rdx}$ dimensionality.

Following the equations in Section 2.2.2 for the closed-loop adaptive DPD, the new reduced set of transformed DPD coefficients $\hat{\boldsymbol{w}}$ at the k^{th} training iteration will be calculated as

$$\hat{\boldsymbol{w}}^{k+1} = \hat{\boldsymbol{w}}^k + \mu\left(\hat{\boldsymbol{U}}^H\hat{\boldsymbol{U}}\right)^{-1}\hat{\boldsymbol{U}}^H\boldsymbol{e} \tag{3.15}$$

and the (full-length) forward path DPD coefficients will be obtained through back transformation of the feedback path coefficient set as

$$\boldsymbol{w} = \boldsymbol{V_{rdx}}\hat{\boldsymbol{w}}. \tag{3.16}$$

One important remark is that the PCA transformation matrix needs only to be calculated once (i.e., can be pre-calculated offline) and will be valid as long as the statistical properties of the input data or the PA operating conditions do not vary significantly. Such robustness is demonstrated in [Pha19a]. Nevertheless, if the recalculation of the PCA transformation matrix is required due to highly diversified PA operating conditions and input waveforms, the computational cost would not be negligible, as it would imply 1) calculating the covariance matrix with computational $\mathcal{O}\left(LN^2\right)$ and 2) generating the eigenvalue decomposition of the covariance matrix with computational cost $\mathcal{O}\left(N^3\right)$. Thus, the total computational cost of the PCA is $\mathcal{O}\left(LN^2 + N^3\right)$ [Sha07]. If the basis functions in matrix $\boldsymbol{U}$ needed to be frequently updated, the corresponding update in the PCA transformation matrix could be impractical in embedded processors with limited capabilities. To handle such challenge, an independent DPD parameter estimation method based on the adaptive PCA (APCA) algorithm is presented in Section 3.2.

Partial Least Squares (PLS)

The partial least squares (PLS) technique is another statistical method used to build a new basis of components from linear combinations of the original basis functions. While PCA obtains new components maximizing their ow variance, PLS creates its composite variables by maximizing the covariance between such new components and a reference signal (i.e., the PA output signal $\boldsymbol{y}$) within the context of linear regression. Thanks to this, PLS is able to outperform PCA when reducing the order of PA behavioral modeling or DPD linearization. As demonstrated in [Gil16, Pha18c], PLS can guarantee well-conditioned identification when reducing the number

of estimated parameters without loss of accuracy, in contrast to PCA, and the complexity of the Moore-Penrose matrix inversion operation when applying LS is remarkably reduced.

Proposed in [Jon93], the iterative SIMPLS algorithm can be used to get a new basis of orthonormal components with PLS. Considering the same variables notation deployed for the PCA algorithm and (3.15), for PLS we have that the transformed coefficients at the k^{th} training iteration will now be calculated as

$$\hat{\boldsymbol{w}}^{k+1} = \hat{\boldsymbol{w}}^{k} + \mu \hat{\boldsymbol{U}}^{H} \boldsymbol{e}, \tag{3.17}$$

due to the orthonormal property of the transformed matrix $\hat{\boldsymbol{U}}$, where $\hat{\boldsymbol{U}}^{H}\hat{\boldsymbol{U}} = \mathbf{I}$. Therefore, the updating of the transformed coefficients is considerably simplified. As per (3.16), the anti-transformed forward path DPD coefficients $\boldsymbol{w}$ will be calculated multiplying the PLS-transformed coefficients $\hat{\boldsymbol{w}}$ by the $N \times N_{rdx}$ dimensional transformation matrix $\boldsymbol{V_{rdx}}$. Thanks to considering the PA output signal $\boldsymbol{y}$ when creating the transformation matrix, PLS will be more robust than PCA in terms of performance degradation when the number of transformed basis and thus transformed DPD coefficients are reduced during identification (at feedback path).

Another more computationally efficient variant of the PLS algorithm for DPD application is presented in [Pha19a]. While in PLS the number of transformed basis is kept constant over the DPD training process, in this approach a dynamic PLS (DPLS) algorithm is employed to apply variable basis reduction depending on the residual linearization error metric (i.e., difference between the desired PA output signal and the one that is measured). Therefore, the DPLS technique allows efficiently adjusting the number of coefficients required to meet the linearity goal. Such DPLS DPD adaptation procedure is summarized in the flowchart shown in Fig. 3.19:

To generate the reduced transformation matrix, now a SIMPLS modification called dynamic orthonormal transformation matrix (DOTM) algorithm is proposed . As fully detailed in [Pha19a], in the DOTM algorithm the number of columns of $\boldsymbol{V_{rdx}}$ is iteratively increased and calculated until the power of the estimated error is close enough to a desired threshold which is a percentage δ defined over the power of the error signal $\boldsymbol{e}$. At every DPD training iteration k, the number of columns N_{rdx} (with $N_{rdx} < N$) of the reduced transformed matrix varies to account for the minimum number of columns that meet the previous optimization goal. Again, the transformed coefficients increment at every iteration will be calculated as $\boldsymbol{\Delta}\hat{\boldsymbol{w}} = \mu \hat{\boldsymbol{U}}^{H} \boldsymbol{e}$ according to (3.17), and the forward path full set of DPD coefficients will be obtained by anti-transforming the DPLS transformed reduced set of DPD coefficients over as many DPD training iterations as needed to meet, for instance, an ACPR level goal.

Finally, in [Pha19b] the authors combine PCA and DPLS, which is equivalent to the canonical correlation analysis (CCA) updating solution, that goes beyond DPLS towards further reducing the amount of transformed DPD coefficients during estimation while meeting the same linearization performance.

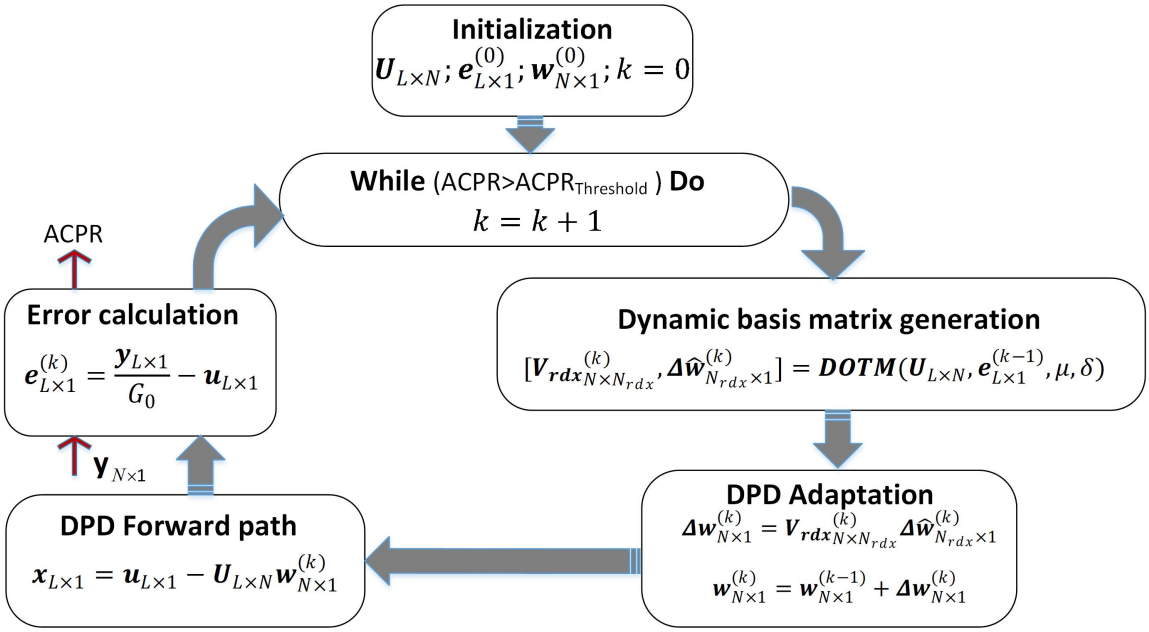


Figure 3.19: Flowchart of the DPD estimation/adaptation using dynamic basis matrix [Pha19a].

Feature Extraction Use Case

Combining CFR and PCA-assisted DPD for efficient Wireless Backhauling

This Use Case shows the application of the previously introduced PCA technique, as DPD order reduction method in the feedback path, and evaluates the linearity performance in a wide-band wireless backhaul transmitter when applying a five-stage PC-CFR (as per Section 2.3) and a direct learning DPD with memory effects compensation (combined as shown in Section 2.2). The goal is to enhance the transmitted output power, and thus the power efficiency of the overall system, while fulfilling the communications system quality requirements and regulations.

The experimental testbed that was deployed for evaluating the PC and DPD linearization performance is illustrated in Fig. 3.20 [LB14]. A four contiguous channel aggregation configuration was deployed considering 28-MHz bandwidth and 1024-QAM modulation signals (4CH1011 -one channel is off), featuring 112 MHz bandwidth and 12-dB PAPR in total. The generated waveform passes through the PC and DPD blocks implemented in MATLAB and is downloaded for playback in the Texas Instruments boards (TSW1400EVM pattern generator and TSW30H84EVM DACs and IQ modulators) that output the signal that will be fed into the GaN PA (DUT) with 16-bit resolution at 625 Msa/s after upconversion with an IQ modulator to the 2 GHz DUT input IF. The DUT is a 6 GHz RF Unit used for microwave backhaul applications with a PA based in a 15 Watts GaN transistor. The output of the PA is digitized using a DSO at 20 Gsa/s.

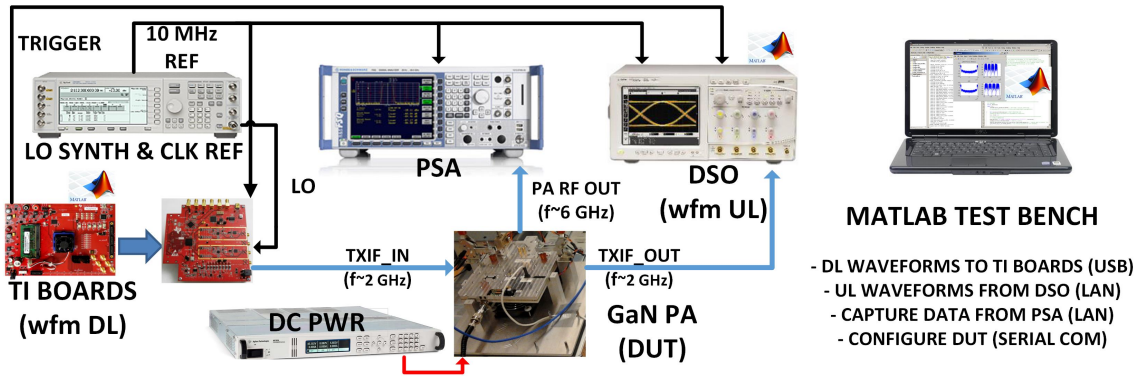


Figure 3.20: Wideband CFR and DPD test bench.

Fig. 3.21-left shows the linearity performance in terms of NMSE (worst case channel), ACPR and EVM (worst channel case), when considering different PAPR reduction factors and with DPD. By combining the PC technique and DPD, and considering an EVM threshold of 3%, we can observe that the maximum allowed PAPR reduction is around 2.75 dB (without DPD would be around 2 dB). Moreover, with PC and keeping the same DPD gain, the ACPR is slightly improved (around 1 dB) since we have prevented the signal peaks from going into hard compression.

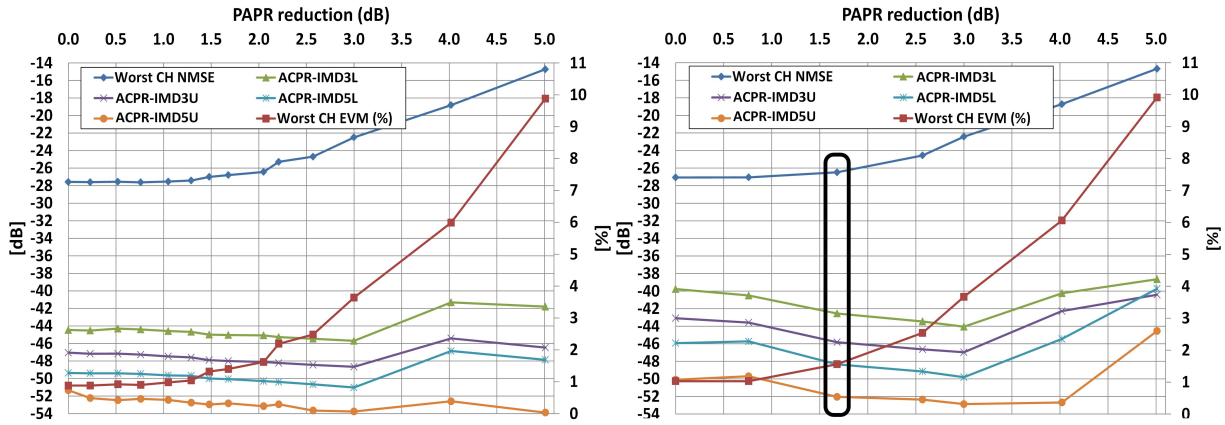


Figure 3.21: NMSE, EVM and ACPR vs PAPR reduction with DPD at nominal output power (left), and at +2 dB output power (right).

Alternatively, thanks to the PAPR reduction the mean output power could be increased without excessively compromising the ACPR and EVM figures. Fig. 3.21-right shows again the linearity performance for different PAPR reduction factors but now at 2 dB higher output power (thus driving the PA further into compression). Severe PAPR reductions imply that the switched-off channel will be more interfered by the neighboring channels. However, there are PAPR reduction values that, when applying PC+DPD, can keep the EVM below 2% despite the 2 dB power increase. In addition, in comparison to the previous nominal power case with DPD, the ACPR is degraded less than 2 dB (see Fig. 3.21-right and Fig. 3.23).

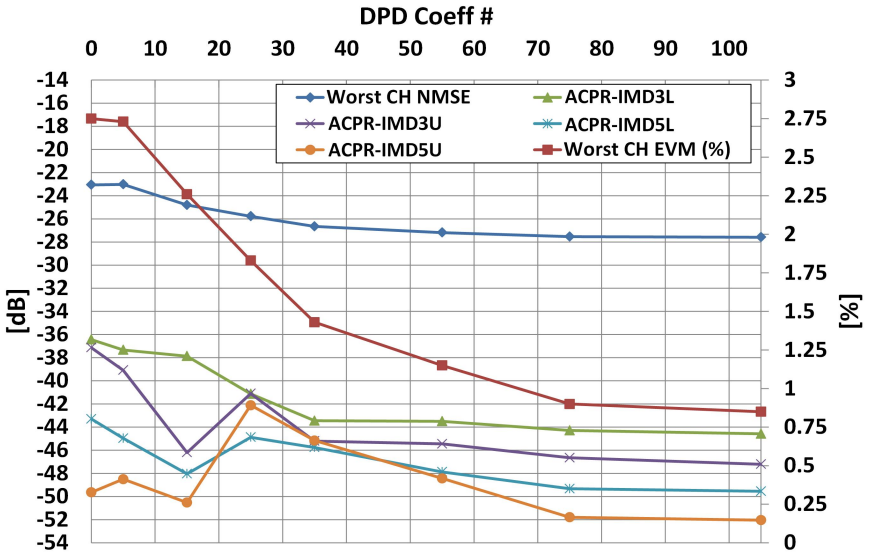


Figure 3.22: NMSE, EVM and ACPR vs number of DPD coefficients.

Fig. 3.22 shows the DPD linearity performance for different memory lengths and thus number of coefficients. As it can be observed, after a certain value of coefficients (e.g., 75) there is no further improvement and, moreover, increasing the number of coefficients could only lead to worsen the linearity figures due to producing an ill-conditioned system. As listed in Table 3.1, thanks to the PCA technique it is possible to apply model order reduction and, with the same number of coefficients, obtaining better in-band linearity figures. However, as shown in Fig. 7, the out-of-band linearity is degraded when increasing the model order reduction factors. As expected, there is a trade-off between the ACPR performance and the DPD complexity in terms of number of coefficients.

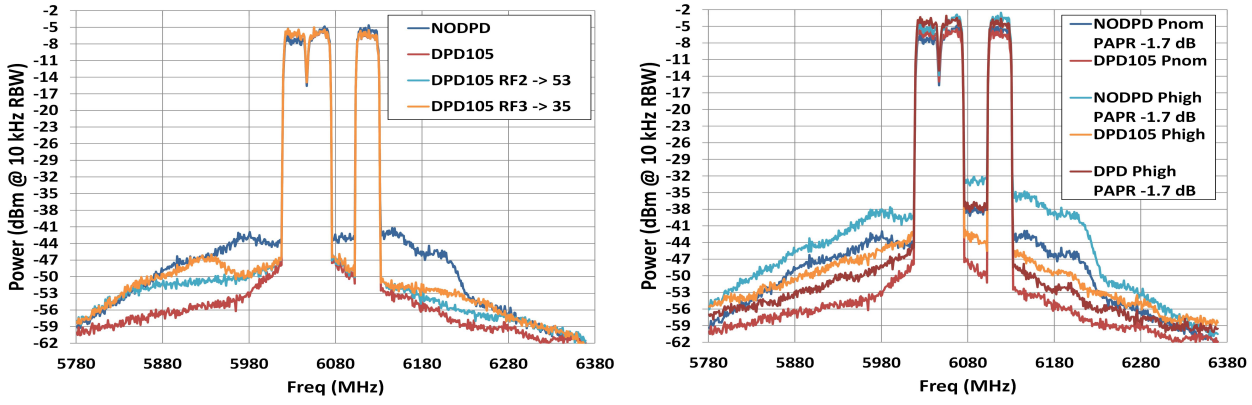


Figure 3.23: Spectrum plots of a 112 MHz BW multi-channel 1024-QAM.

The use of the PC technique combined with DPD with model order reduction, helps to mitigate the trade-off between linearity, efficiency and computational complexity. We have experimentally demonstrated that gaining near 2 dB of output power thanks to the use of DPD

Table 3.1: DPD linearization performance.

Configuration	ACPR (dBc)				Worst Channel EVM (%)	Worst Channel NMSE (dB)	No. Coeff.
	IMD3 L	IMD3 U	IMD5 L	IMD5 U			
No DPD	-36.4	-37.1	-43.3	-49.6	2.75	-19.05	-
DPD 3 taps	-37.9	-46.2	-48.0	-50.5	2.26	-25.57	15
DPD 21 taps	-44.6	-47.2	-49.6	-52.0	0.85	-28.44	105
DPD 21 taps RDF=2	-41.8	-45.6	-44.7	-50.2	0.99	-28.11	53
DPD 21 taps RDF=3	-39.9	-44.4	-43.6	-49.6	0.92	-28.09	35
DPD 21 taps RDF=7	-37.2	-41.4	-42.6	-49.9	1.24	-27.21	15

and CFR techniques can be achieved minimizing the computational resources with the PCA technique and still keeping reasonable performance.

3.2 Independent Parameter Estimation for SISO Polynomial Digital Predistorters

3.2.1 Introduction to implementation-friendly DPD estimation

While the forward path of the DPD must operate in real-time in the FPGA programmable logic (PL), the adaptation of the DPD coefficients in 4G radios could be done typically in a processing system (PS) in a much slower time scale than in the forward path (i.e., not in real time). When looking at the feedback path, in order to deal with an overdetermined system or an ill-conditioned covariance matrix, most of the times the straightforward solution found in literature consists in using MATLAB's backslash operator (otherwise known as the "mldivide" function) that employs a QR solver for dense non-square matrices like those typically found in DPD. Either in commercial products or in publications addressing FPGA implementation, one of the most common solutions to avoid calculating the inverse of the covariance matrix in (2.19) is extracting the parameters through QR factorization combined with recursive least squares (QR-RLS) [Mur06]. In [Wan17a], the authors present a method to reduce the computational complexity of the direct learning architecture by making use of the constant covariance matrix to avoid recalculating the time-varying matrix for OFDM-like input signals seen as stationary ergodic random processes. The approach and the complexity versus convergence benchmarking are worth considering but, in this contribution, we take benefit from the independence of the transformed basis in the feedback path to propose an online scalable solution to meet the requirements at the lowest usable complexity by using a conveniently modified PCA variant able to overcome the drawbacks in terms of computational complexity and training time stated in Section 3.1.4 for limited capacity processors.

In [Pha18a], we proposed the parameter extraction of the DPD behavioral models using a new method based on the adaptive principal component analysis (APCA) technique. The APCA approach resembles the SVD method but the main difference relies on the fact that APCA is able to continuously track and adapt to the evolution of the eigenvectors required for doing DPD [Hay09]. The proposed block deflated APCA (BD-APCA) technique is a modification of the complex domain generalized Hebbian algorithm (CGHA) [Zha97]. In this section, it will be proven that, thanks to the APCA technique, we can iteratively find a new orthogonal basis that will solve the covariance matrix inversion problem by converting the original least squares regression into a set of independent adaptive least mean square (LMS) identifications. This process can run online in an FPGA for not only ensuring a proper well-conditioned estimation, but also allowing to reduce the number of parameters in the identification process. Unlike in [Gil13], we do not assume that the transformation matrix is calculated only once, offline, and then used for all the data having the same statistical properties as the one used to extract the transformation matrix. Instead, a step further is taken towards a possible online implementation of the APCA by demonstrating (both theoretically and experimentally) the benefits of the independent calculation of the DPD coefficients. Alleviating the data processing implementation bottlenecks of the PCA technique, the scalable APCA method is in this section further experimentally validated for DPD application by employing an increasing transformation matrix which produces an increasing set of independently calculated coefficients only if the system performance requirements are not yet met (i.e., starting with one coefficient, and increasing the number of coefficients by one at every iteration).

In Section 3.2.2, the independent identification/adaptation subsystem based on the APCA method to further reduce the number of the parameters required for a robust identification is presented. Section 3.2.3 describes the experimental test bench and shows the experimental results proving the advantages of the proposed independent DPD estimation method using the APCA algorithm.

3.2.2 Independent DPD parameter estimation based on the APCA algorithm

As discussed previously, there are several order reduction techniques published in literature targeting a simplified implementation of the DPD function in the forward path. In this section instead, we will focus on how to address the LS matrix inversion of the covariance matrix to extract the DPD model parameters in order to enable implementation in an FPGA (containing a programmable logic device and a processing system).

Following with the mathematical developments detailed for PCA in Section 3.1.4, by taking into account the orthogonal basis functions in $\hat{\boldsymbol{U}}$, we have that

$$\left(\hat{\boldsymbol{U}}^H\hat{\boldsymbol{U}}\right)^{-1} = \left(\lambda_1^{-1}, \cdots, \lambda_j^{-1} \cdots, \lambda_{N_{rdx}}^{-1}\right)\boldsymbol{I} \tag{3.18}$$

with λ_j being the eigenvalues of $\boldsymbol{U}^H\boldsymbol{U}$ and $\boldsymbol{U}\boldsymbol{U}^H$ (with $j = 1, \cdots N_{rdx}$). The coefficients can be now estimated independently in an LMS fashion at every sample iteration n, and thus (3.15) becomes

$$\begin{pmatrix} \hat{w}_1[n+1] \\ \vdots \\ \hat{w}_j[n+1] \\ \vdots \\ \hat{w}_{N_{rdx}}[n+1] \end{pmatrix} = \begin{pmatrix} \hat{w}_1[n] \\ \vdots \\ \hat{w}_j[n] \\ \vdots \\ \hat{w}_{N_{rdx}}[n] \end{pmatrix} + \mu \begin{pmatrix} \lambda_1^{-1}\psi_1[n] \\ \vdots \\ \lambda_j^{-1}\psi_j[n] \\ \vdots \\ \lambda_{N_{rdx}}^{-1}\psi_{N_{rdx}}[n] \end{pmatrix} e[n] \tag{3.19}$$

By exploiting the orthogonality of the resulting transformed basis functions, the coefficient adaptation can therefore be carried out independently as follows,

$$\hat{w}_j[n+1] = \hat{w}_j[n] + \mu\lambda_j^{-1}\psi_j[n]e[n] \tag{3.20}$$

with $j = 1, \cdots, N_{rdx}$ and where $\psi_j[n]$ is the j^{th} transformed basis function at time n. A schematic flowchart describing the independent DPD extraction is depicted in Fig. 3.24.

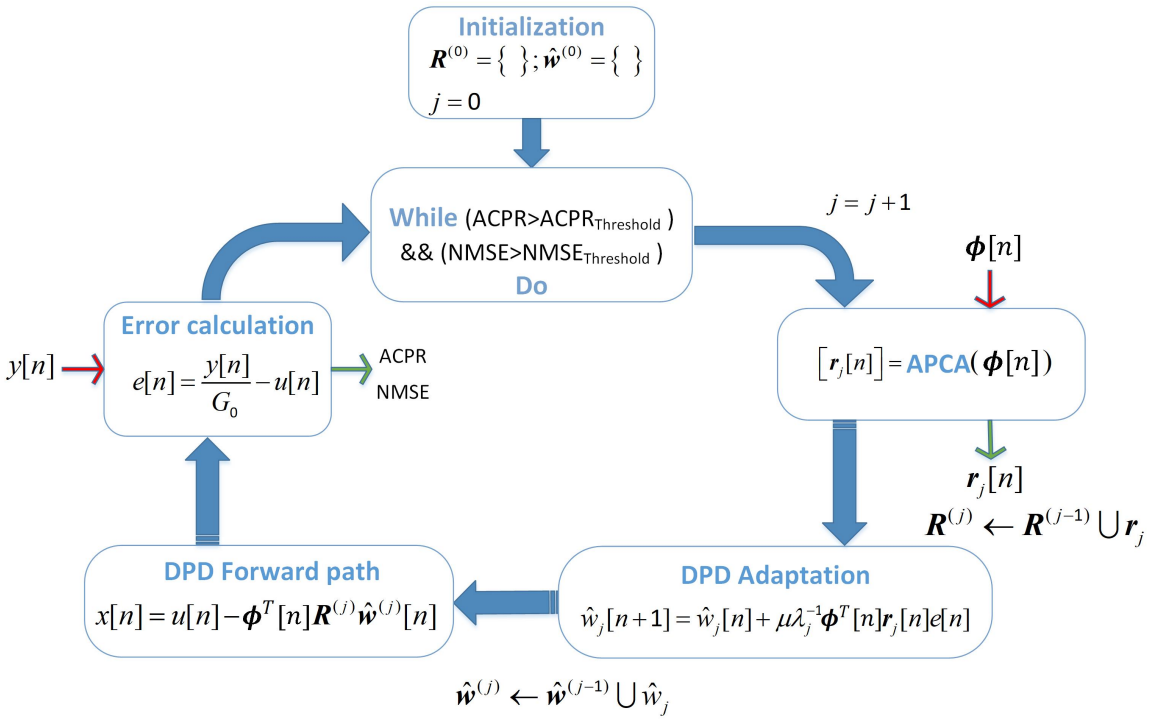


Figure 3.24: Flowchart of the independent DPD identification process using APCA.

The goal is to estimate the minimum necessary number of transformed coefficients $\hat{w}_j$ necessary to meet the target linearity levels, specified in terms of ACPR and NMSE. As explained in [Pha18a], with the proposed block deflated adaptive principal component (BD-APCA) algorithm, the columns $\boldsymbol{r}_j$ ($j = 1, 2, \cdots, N_{rdx}$) of the transformation matrix $\boldsymbol{R}$ are iteratively found one by one. Therefore, the next column is estimated by using the values of the previously extracted

components. Eventually, the $N \times N_{rdx}$ transformation matrix $\boldsymbol{R} = (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_j, \cdots, \boldsymbol{r}_{N_{rdx}})$ will converge to $\boldsymbol{V}$. The BD-APCA algorithm is detailed as follows:

BD-APCA procedure

Input: The $N \times 1$ vector containing the basis functions $\boldsymbol{\phi}^T[n] = \Big(\varphi_1[n], \cdots, \varphi_i[n], \cdots, \varphi_N[n]\Big)$ from matrix $\boldsymbol{U}$.

Output: The transformation matrix $\boldsymbol{R} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_{N_{rdx}})$.

1. Set $j = 1$, initialize small random values to the matrix $\boldsymbol{R}$ and assign $\boldsymbol{b}_j[n] = \boldsymbol{\phi}^*[n]$, with $n = 0, 1, \cdots, L-1$.
   where * denotes the complex conjugate.
2. Set $k = 1$, $\boldsymbol{B}_j = (\boldsymbol{b}_j[0], \boldsymbol{b}_j[1], \cdots, \boldsymbol{b}_j[L-1])^T$.
   2.1) For $n = 0, 1, \ldots, L-1$, update the vector $\boldsymbol{r}_j$ as follows,

$$\boldsymbol{r}_{j,k}[n+1] = \boldsymbol{r}_{j,k}[n] + \eta_{j,k}(d_{j,k}[n])^* \big(\boldsymbol{b}_j[n] - d_{j,k}[n]\boldsymbol{r}_{j,k}[n]\big) \tag{3.21}$$

   where $d_{j,k}[n] = \boldsymbol{r}_{j,k}^H[n]\boldsymbol{b}_j[n]$ and the learning-rate parameter is

$$\eta_{j,k} = \sigma \cdot trace(\boldsymbol{B}_j^H \boldsymbol{B}_j)/k \tag{3.22}$$

   with σ being a constant factor. The trace can be calculated as the sum of the vector's squared modulus, thus avoiding the need to calculate the full correlation matrix.

   2.2) Increment k by 1, go back to step 2.1, and continue until the values of $\boldsymbol{r}_{j,k}$ become steady. Store the estimated eigenvector $\boldsymbol{r}_j = \boldsymbol{r}_{j,k}[L]$. Go to step 3.
3. Deflat the data vectors $\boldsymbol{b}_{j+1}[n]$, with $n = 0, 1, \cdots, L-1$.

$$\boldsymbol{b}_{j+1}[n] = \boldsymbol{b}_j[n] - \big(\boldsymbol{r}_j^H \boldsymbol{b}_j[n]\big)\boldsymbol{r}_j \tag{3.23}$$

4. Increment j by 1 and go back to step 2, continue until the last $\boldsymbol{r}_j$ is obtained.

Therefore, as shown in Fig. 3.24, until the desired linearity levels are met, the algorithm increases the number of transformed coefficients to be estimated. After several APCA iterations, a new transformation vector $\boldsymbol{r}_j$ is first obtained. This vector, together with the other $j-1$ previously calculated vectors (i.e., $j-1, j-2, \cdots, 1$), define the j components of the transformation matrix $\boldsymbol{R}^{(j)}$. Then, each one of the j transformed coefficients $\hat{w}_j[n]$ can be estimated/updated independently (e.g., in parallel) by following an LMS approach,

$$\hat{w}_j[n+1] = \hat{w}_j[n] + \mu\lambda_j^{-1}\boldsymbol{\phi}^T[n]\boldsymbol{r}_j[n]e[n] \tag{3.24}$$

Consequently, the input-output relationship of the DPD function in the forward path described in (2.17) can be rewritten as

$$x[n] = u[n] - \boldsymbol{\Phi}^T[n]\hat{\boldsymbol{w}}[n] \tag{3.25}$$

or, alternatively, be expressed in terms of the original basis functions $\boldsymbol{\varphi}[n]$ and the transformation matrix $\boldsymbol{R}$.

$$x[n] = u[n] - \boldsymbol{\phi}^T[n]\boldsymbol{R}\hat{\boldsymbol{w}}[n] \tag{3.26}$$

In the following section, experimental results proving the robustness of the independent DPD algorithm described in Fig. 3.24 will be shown.

3.2.3 Experimental Results

The proposed independent DPD parameter extraction using the APCA technique was experimentally evaluated with a 5G-like communication system. Given the unavailability of 5G-NR waveforms at the time of testing, we employed a carrier-aggregated fast-convolution filter bank multi-carrier (FC-FBCM) signal with 80 MHz bandwidth built with four 64-QAM modulated 20 MHz FC-FBMC channels that were carrier aggregated. Each channel follows an OFDM-like waveform configuration and a frame structure similar to that used for LTE FDD (Rel. 9) and enables spectral coexistence with LTE primary systems through subcarrier group deactivation. These complex waveforms feature around 13 dB PAPR. In our experiments, we used a MATLAB-controlled digital linearization test bench (shown in Fig. 3.25) interfacing some commercial boards for waveform playback and data capture, digital-to-analog conversion (DAC), IQ modulation and analog-to-digital conversion (ADC) for direct RF sampling (i.e., TI TSW1400EVM and TSW30H84EVM at Tx side and TI ADC32RF45EVM and TSW14J56EVM at Rx side). In order to account for the out-of-band distortion, a 368.64 MSa/s DPD signal with 240 MHz bandwidth was up-converted to the 875 MHz RF frequency to feed a class-J PA based on the Cree CGH35030F GaN HEMT. The PA output signal (with +28 dBm mean output power) was attenuated, RF sampled at 2457.6 MSa/s, and further downsampled to the DPD signal sample rate for time-alignment and DPD processing. An Agilent E4440A spectrum analyzer was used to characterize the spectrum and ACPR at the output of the PA, three Agilent E4438C signal generators were used as data converter clocks and IQ modulator local oscillator source (all sharing the 10 MHz clock reference), and an Agilent N6710B modular dc power system was employed to supply the active RF elements of the test setup.

The independent DPD closed-loop adaptation described in Fig. 3.24 has been validated in a MATLAB-controlled hardware test bench. Thus, instead of running the algorithm sample by sample (as it would be done in an FPGA) we considered the use of input-output data vectors for faster evaluation of the proposed algorithms. At every iteration a different set of data was used,

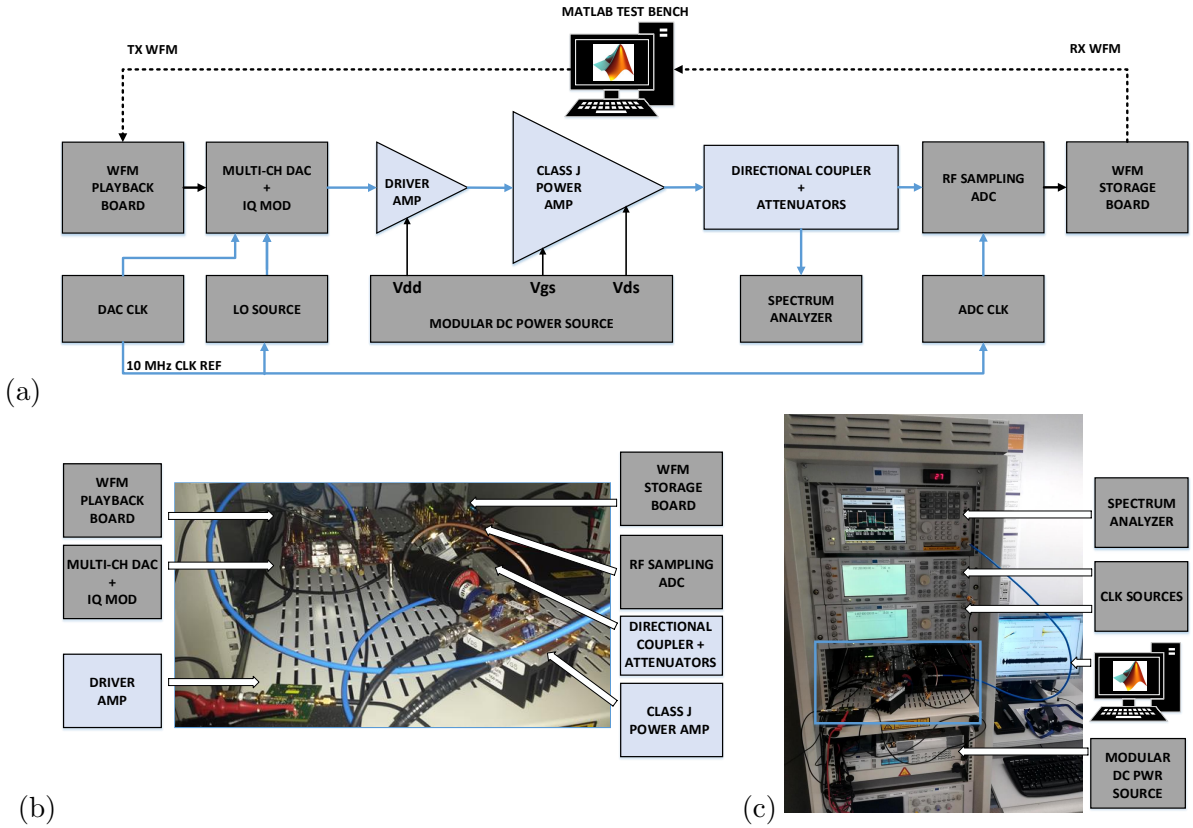


Figure 3.25: (a) Block diagram of the test setup employed for experimental validation. The digital linearization test bench combines (b) waveform generation and capture boards, RF analog control parts including the PA and (c) laboratory instrumentation. TX: transmitted; RX: received; WFM: waveform; CH: channel; MOD: modulator; AMP: amplifier; ADC: analog-to-digital converter; DAC: digital-to-analog converter; CLK: clock; LO: local oscillator; DC PWR: dc power; Vdd: drain-to-drain voltage ; Vgs: gate-to-source voltage; Vds: drain-to-source voltage.

presenting different PAPR values (around 13 $\pm$ 1 dB) to prove the robustness of the proposed DPD linearization algorithm.

As shown in (2.19), typically, the Moore-Penrose inverse i.e., $(\boldsymbol{U}^H\boldsymbol{U})^{-1}\boldsymbol{U}^H$, is used to solve the LS identification. However, when the covariance matrix $(\boldsymbol{U}^H\boldsymbol{U})$ is ill-conditioned, the values of the estimated coefficients are no longer reliable. Table 3.2 shows the linearization achieved (in terms of ACPR, NMSE and EVM in the worst-case channel) when considering two cases: a) the direct estimation of the best performing 322 coefficients found by brute force using MATLAB's backslash operator (since using the Moore-Penrose inverse the coefficient estimation diverges) and b) the independent DPD adaptation of 60 coefficients using APCA.

The original data matrix containing 322 basis functions could have been reduced by applying some model order reduction such as OMP. Nevertheless, this is not the scope of this contribution and, more important, by assuming an initial data matrix with 322 basis functions we can prove

Table 3.2: DPD Performance Comparison

Configuration 80 MHz FC-FBMC	Coeff. No.	NMSE [dB]	ACPR [dBc]	EVM [%]
No DPD	-	-18.5	-36.25	5.75
Full DPD w/ MATLAB's '\'	322	-41.1	-50.7	1.05
Indep. DPD w/ APCA	60	-38.7	-45.3	1.35

the regularization and coefficient reduction capabilities of the proposed approach. Having said this, and just to have some order reduction assessment, we have applied OMP to reduce the original data matrix in the forward path and found that by using 227 coefficients the performance being attained is very similar to that featured by the 322 coefficients found by brute force. Therefore, the order reduction factor in the feedback path would be around 3.5 if we consider that, to meet the target $ACPR \leq -45$ dBc and $NMSE \leq -38$ dB, only 60 coefficients of the new transformed basis were required. This number is consistent with the experiment results shown in Fig. 3.26 which are further detailed in the next paragraph.

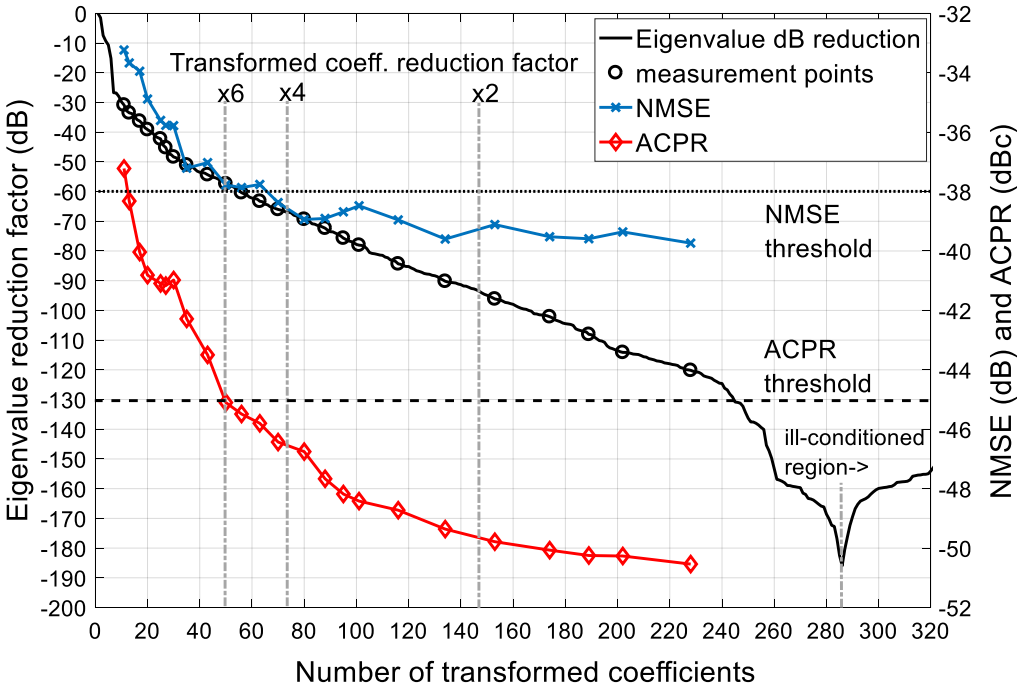


Figure 3.26: NMSE and ACPR vs. coefficients reduction.

Before applying APCA to iteratively build the 60 coefficients, several tests were conducted to characterize the performance of PCA versus the number of transformed coefficients. In order to have good measurement resolution and better exploit the eigenvalue information the following procedure was used: First, the eigenvectors and eigenvalues of the matrix $\boldsymbol{U}^H\boldsymbol{U}$ were calculated

to build the transformation matrix for the full set of 322 coefficients. Secondly, the ratio in dB between the absolute value of the last diagonal element (having the lowest value for a certain number of transformed coefficients) and the absolute value of the first diagonal element (having the maximum value) was characterized in MATLAB as in Fig. 3.26 (black trace). Thirdly, every measurement point defined a maximum number of consecutive transformation matrix columns (i.e., comprising a specific eigenvalue dB reduction factor) and thus DPD coefficients that were employed in every PCA DPD experiment. A total of 25 experiments of 21 iterations each were conducted (different PAPR statistically representative waveforms were applied to check the robustness). After a training period of 10 iterations (by using a decreasing μ weighting factor), the NMSE and ACPR values for every experiment became stabilized and the worst-case value was recorded and annotated in Fig. 3.26. Finally, this information is later used to validate the APCA operation by comparing the number of coefficients progressively obtained by APCA to reach the required performance versus that offered by the original PCA DPD approach taking the whole coefficient set from the first iteration.

By considering the 60 coefficients being required to meet the NMSE and ACPR thresholds and before experimenting with the BD-APCA method, another set of PCA experiments was set to demonstrate the independent estimation of coefficients. In these experiments, at each DPD iteration only a portion of the overall transformed coefficients was updated while the rest was left unaltered. The results in Fig. 3.27 satisfactorily show that similar performance is reached when comparing full update with partial update at different independent update ratios. As expected, a lower number of coefficients being updated takes more iterations to reach the required performance. However, by properly tuning the μ weighting factor being applied at every DPD iteration the delay can be minimized so that the desired performance is reached during the aforementioned 10-iteration training period. It is also worth noting that the order at which we add a new set of independent coefficients will impact on the DPD training time, since each coefficient contributes differently to the linearization performance, but not on the performance itself thanks to the independence of the coefficients.

In the following, we will discuss the APCA experimental validation. Fig. 3.28 shows the NMSE and ACPR evolution when considering that at each iteration a new component is included in the estimation set and thus a new coefficient is estimated. By following the algorithm described in Fig. 3.24, we can see that thanks to the orthogonal property even if at every iteration a new coefficient is included this can be estimated independently and each one contributes to the linearization. After a certain number of coefficients (in our case, 60) the targeted NMSE and ACPR are reached and then we can stop adding components and coefficients to the estimation set. In our experiment we have considered up to 75 iterations to check the method performance stability once the number of coefficients is no longer increased.

Fig. 3.29 shows the evolution of the absolute value of the estimated coefficients. The estima-

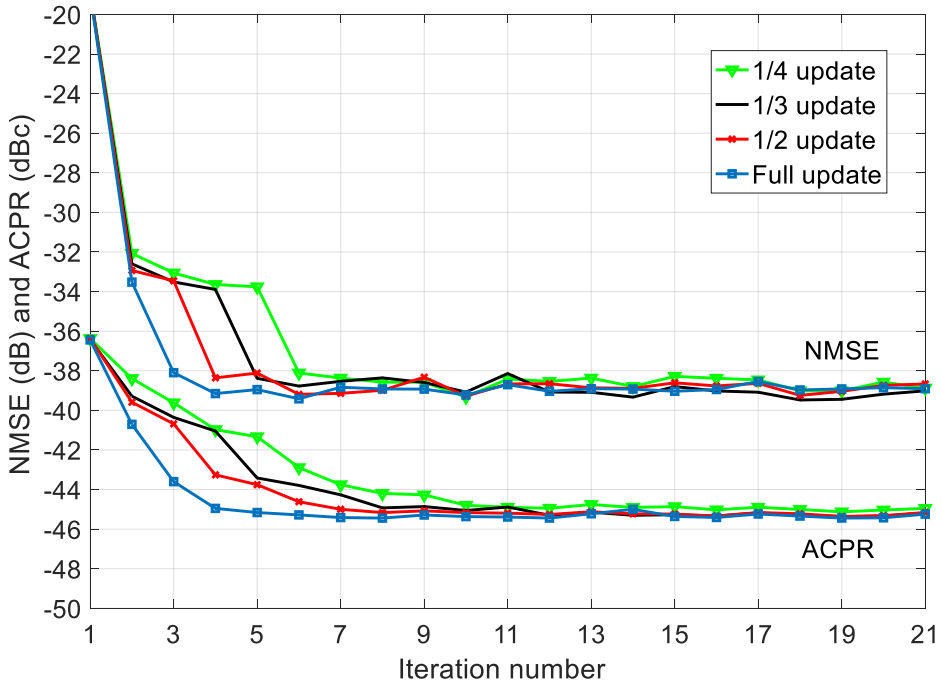


Figure 3.27: Independent partial identification of DPD coefficients per iteration.

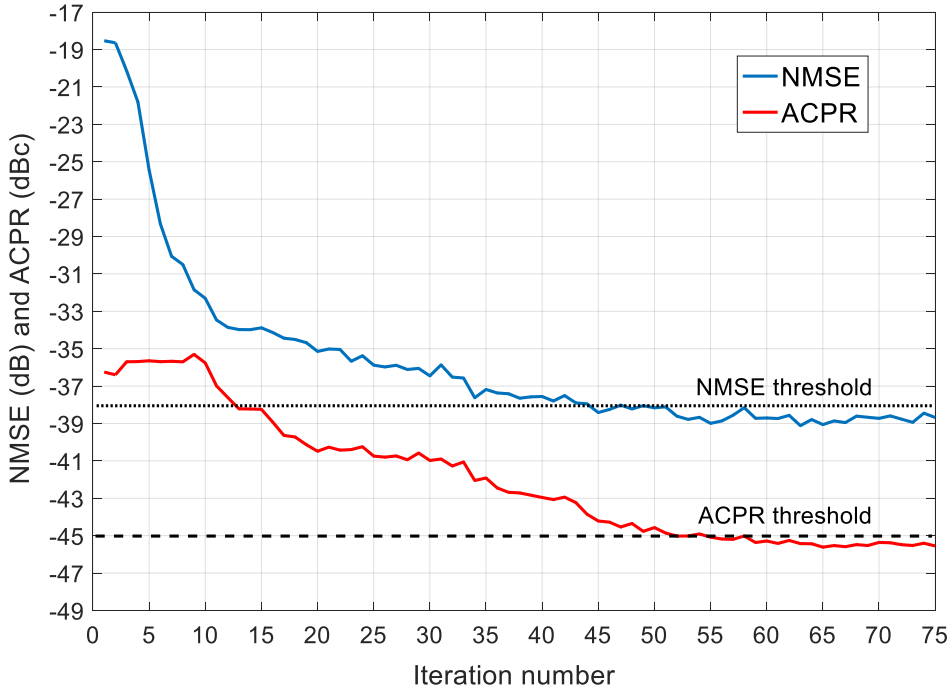


Figure 3.28: Evolution of the NMSE and the ACPR considering up to 60 components and 75 iterations.

tion is perfectly regularized, the absolute value of the coefficients is below 1, and every time a new coefficient is added it converges in few iterations and, because of their independence, the

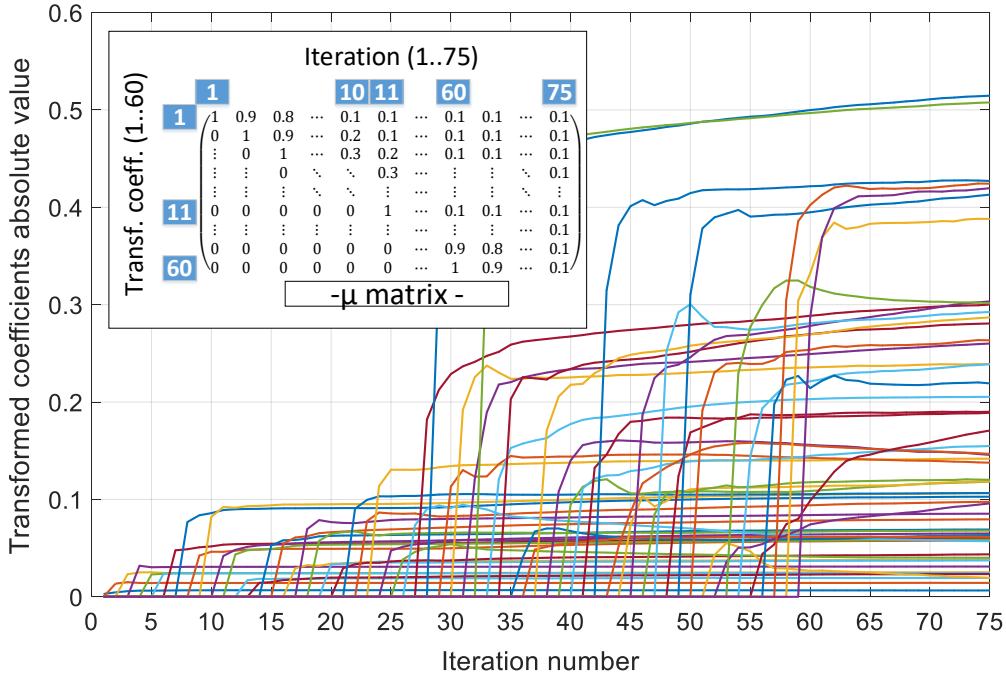


Figure 3.29: Evolution of the absolute value of the 60 DPD coefficients.

value of the already existing coefficients is kept unaltered despite the addition of new coefficients to the estimation set. This cannot be done with the original basis, since the basis functions are not orthogonal and thus the value of the coefficients is dependent on the number of selected basis. This solution is very versatile because thanks to the independent estimation several strategies to reduce the computational complexity can be considered. For example, we could adapt only one coefficient at each iteration following a round robin scheduling and thus leaving the others unadapted until their turn arrives. Fig. 3.29 also depicts the μ DPD weighting factor strategy applied at each experiment iteration. This factor accounts for the first iteration at which such coefficient was generated and the following updates being applied to that coefficient as if there was a specifically bounded training period for it. The rationale behind such strategy is to contribute to guarantee the convergence time and to a shorter extent to help preserving the regularization.

Finally, Fig. 3.30 and Fig. 3.31 show the unlinearized and linearized spectra taken from the spectrum analyzer and the amplitude modulation to amplitude modulation (AM–AM) and amplitude modulation to phase modulation (AM–PM) characteristics respectively, when considering independent DPD adaptation with 60 coefficients.

This contribution has shown the versatility of the independent DPD identification method using the PCA theory to find a transformed matrix with new orthogonal components. With the independent DPD identification, it is possible to tradeoff updating or identification convergence time versus computational complexity. That is, updating in parallel at every iteration (in a more relaxed time scale than the real-time digital predistorter in the forward path), following an LMS

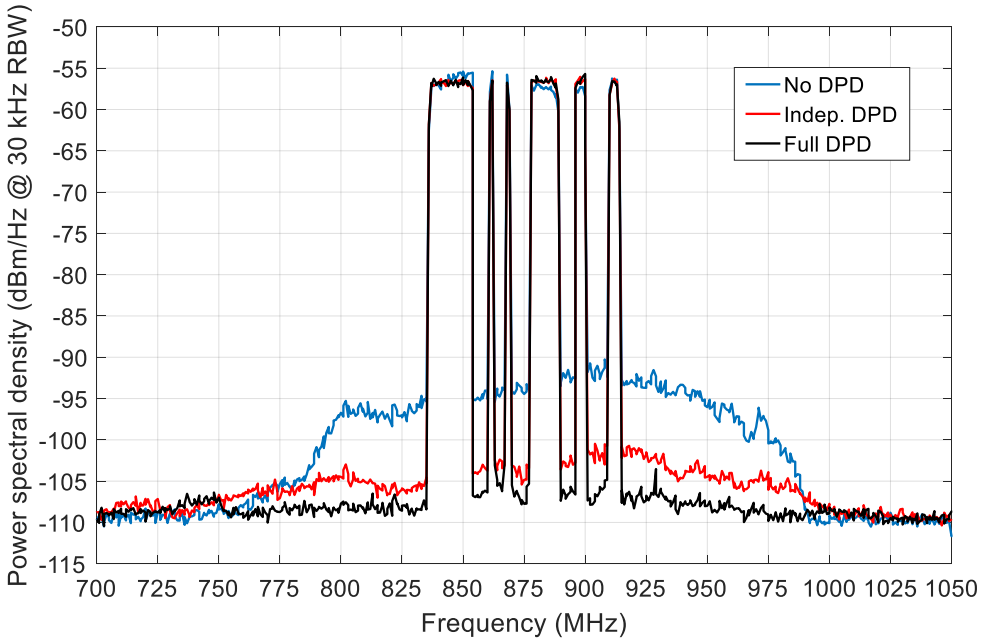


Figure 3.30: Spectra of the PA output before and after DPD linearization.

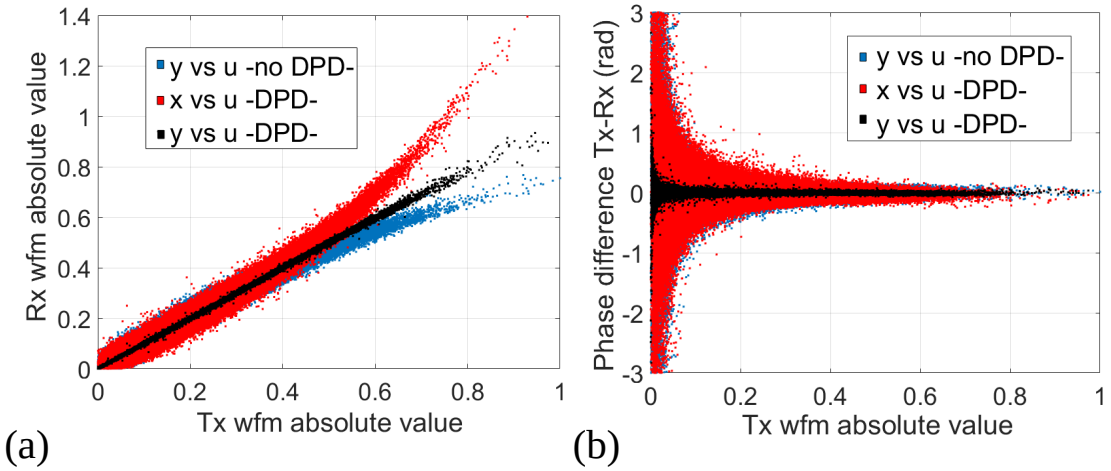


Figure 3.31: AM-AM (a) and AM-PM (b) characteristics before and after DPD linearization.

gradient technique, as many components as you want (tradeoff convergence time vs. FPGA resources). The technique allows guaranteeing robust independent DPD identification using the minimum necessary coefficients in the adaptation subsystem to meet specific linearity levels.

Chapter 4

Machine Learning Techniques for Neural Network SISO Predistorters

4.1 Nonlinear Neural Networks Approaches

4.1.1 Introduction to ANN Topologies and dynamic nonlinear modeling

An artificial neural network (ANN) is a modeling technique, originally inspired by some partial knowledge on the behavior of the neurons in the human brain, that can be trained to learn the structure of the data and model complex nonlinear functions. Essentially, the neurons are distributed between different layers and communicate with each other through neuron output-to-input weighted interconnections (or synapses). Based on the interconnection pattern or architecture we can distinguish between feedforward networks (FFNNs) and recurrent (or feedback) networks (RNNs).

FFNNs, which are among the most used ANN, have unidirectional interconnections between the neurons of every layer since the flow of data is from input to outputs, without feedback (one input pattern produces one output). The most common FFNN is called multilayer perceptron (MLP) which is composed of fully connected layers where all the output activations are composed of a weighted sum of input activations (the neurons of a specific layer are fed by the outputs of all the neurons of the preceding layer). The larger the weight, the more influential the corresponding input will be. Enabling full connection in a densely populated NN may require significant hardware resources but in many applications the weight of some interconnections can be set to zero without loss of accuracy which results in sparsely connected layers. In a RNN, the inputs of the neurons of a specific layer may be fed by the output of the neurons either in the same layer or at any of the following layers which senses time and memory of previous states. The above concepts are shown in Fig. 4.1(a) where a modified MLP-based FFNN is displayed (a classic MLP would have full connection between at all layers).

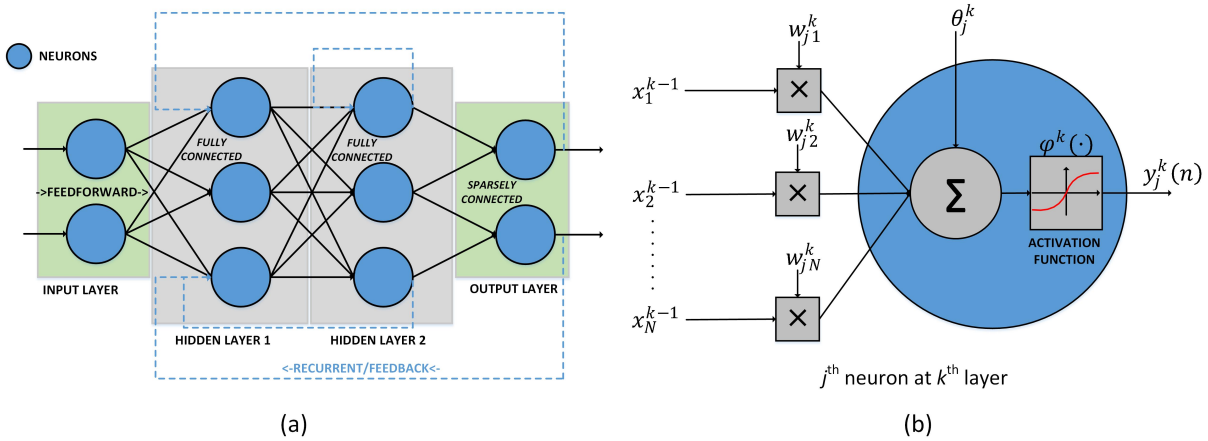


Figure 4.1: (a) FFNN and RRN architectures exemplification and (b) single-layer perceptron (SLP) model.

In more detail, Fig. 4.1(b) shows a single-layer perceptron model and the operation of this fundamental building block of an MLP NN. The j^{th} neuron, of the k^{th} layer receives as input each x_i from the previous layer. Each x_i, with $i = 1, 2, \cdots, N$, is then multiplied by a weight w_{ji} and the resulting values are all summed together. A single bias or offset value θ_j is added to the summation and, finally, an activation or transfer function $\varphi^k(\cdot)$ (different activation functions can be applied to different layers) is applied to provide the output of the j^{th} neuron found in the k^{th} layer, as shown in (4.1):

$$y_j^k(n) = \varphi^k\left(\sum_{i=1}^{N} w_{ji}^k x_i^{k-1}(n) + \theta_j^k\right) \tag{4.1}$$

Some historical brief review follows. When the single-layer perceptron was defined in the late 50's by [Ros58] the activation function being used was the step or threshold function and the concept of hidden layers was not yet exploited. This model was a first practical implementation that could be used for simple linearly-separable binary classification problems but was not valid for more complicated modeling requiring nonlinear outputs. The solution to this limitation came in the mid and late 80's thanks to a few works, such as [Rum86], that considered MLP with hidden layers to enable the NN learning more complicated features, proposed backpropagation algorithms to adjust the weights and minimize the difference between the actual output and the desired output, and employed nonlinear activation functions such as the sigmoid function that could enable gradually changing the weights of the NN and introduce nonlinearity. The universal approximation theorem in [Cyb89] proved that a feedforward ANN with a single hidden layer (a three-layer network considering the input and output layers) and non-constant, bounded, and monotone-increasing continuous activation function can approximate any nonlinear function with any desired error. Fig. 4.2 shows a summary of the main activation functions.

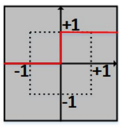
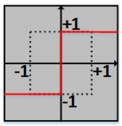
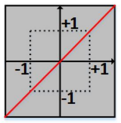
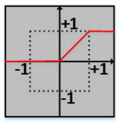
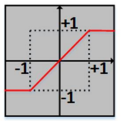
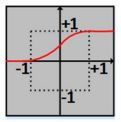
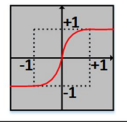
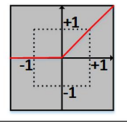
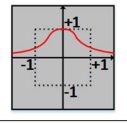
Function name	Equation	Plot
Hard limit (step or threshold)	$\varphi(x) = 0 \; for \; x < 0$ $\varphi(x) = 1 \; for \; x \geq 0$	
Symmetrical hard limit	$\varphi(x) = -1 \; for \; x < 0$ $\varphi(x) = 1 \; for \; x \geq 0$	
Linear	$\varphi(x) = x$	
Saturating Linear	$\varphi(x) = 0 \; for \; x < 0$ $\varphi(x) = x \; for \; 0 \leq x \leq 1$ $\varphi(x) = 1 \; for \; x > 1$	
Symmetric Saturating Linear	$\varphi(x) = -1 \; for \; x < -1$ $\varphi(x) = x \; for \; -1 \leq x \leq 1$ $\varphi(x) = 1 \; for \; x > 1$	
Log-Sigmoid	$\varphi(x) = \frac{1}{1 + e^{-x}}$	
Hyperbolic Tangent Sigmoid	$\varphi(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$	
Positive Linear (a.k.a Rectified Linear Unit or ReLU)	$\varphi(x) = 0 \; for \; x < 0$ $\varphi(x) = x \; for \; x \geq 0$	
Gaussian (multi-variate with variance σ_i^2, t_i is the ith centre of the Green's function)	$\varphi(x, t_i) = G(\|\|x - t_i\|\|) =$ $= exp(-\frac{1}{2\sigma_i^2}\|\|x - t_i\|\|^2)$	

Figure 4.2: ANN activation or transfer functions.

In the past, and generally speaking, FFNs were considered static and memoryless in the sense that the response of an input was independent of the previous network state, while RNNs were considered dynamic systems because of the feedback connections. Nowadays, and given the highest complexity of RNNs versus FFNNs, the RNN architectures are frequently unrolled in a way that are re-drawn and re-formulated similarly as a FFNN to simplify the processing complexity. In addition, the need for modeling nonlinear system dynamics considering memory effects has grown over the last decades in multiple applications. For instance, the ever-increasing signal bandwidth at each wireless communication standard generation makes the modeling of PA memory effects, which are more evident when this component is excited by higher bandwidth

signals, a relevant topic to be accounted to enhance the performance of the physical layer. In order to solve time series prediction and thus enable dynamic nonlinear system identification, the focused time-delayed neural networks (FTDNNs), which include tapped delay lines to generate delayed samples of the input variables, have been proposed. The FTDNN can be seen as combining a linear time invariant (LTI) system such as a finite impulse response (FIR) filter, which enables performing dynamic mappings depending on past input values, and a nonlinear memoryless MLP network that can be trained using static backpropagation algorithms. The memory depth of the system being modeled will be reflected on the length of the taps imposed by the required bandwidth accuracy. A FTDNN structure could be seen as a special case of the Wiener model (i.e., a linear time-invariant system followed by a memoryless nonlinear system). Fig. 4.3 shows a four-layer architecture (with two hidden layers) of a fully connected FTDNN whose input-output relation is defined in (4.2) according to the notation of the aforementioned SLP concept.

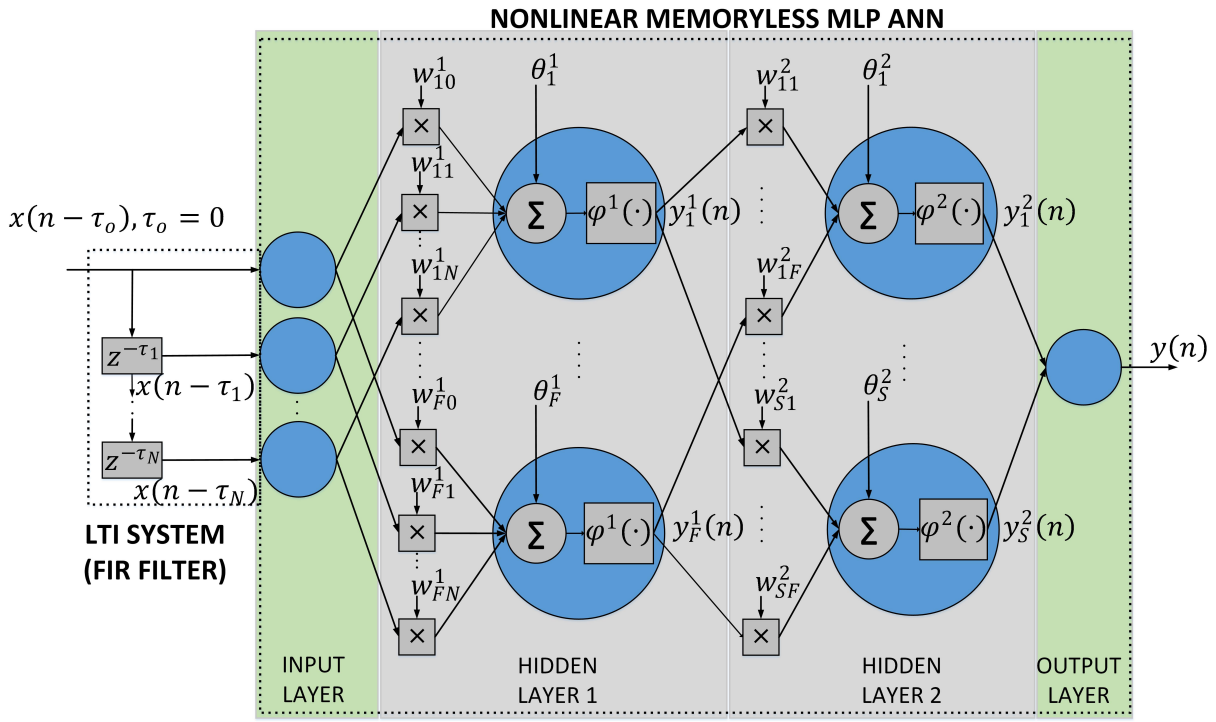


Figure 4.3: Four-layer FTDNN architecture.

$$y(n) = \sum_{k=1}^{S} \varphi^2 \left(\sum_{j=1}^{F} w_{kj}^2 y_j^1(n) + \theta_k^2 \right) = \sum_{k=1}^{S} \varphi^2 \left(\sum_{j=1}^{F} w_{kj}^2 \varphi^1 \left(\sum_{i=0}^{N} w_{ji}^1 x_i(n-\tau_i) + \theta_j^1 \right) + \theta_k^2 \right) \tag{4.2}$$

In this example, the input layer contains $N+1$ neurons (including the input signal and all the delayed versions, z^{-1} is the unit delay operator), the first hidden layer has F neurons, the second hidden layer has S neurons and there is a final output layer with a single neuron. For the sake of simplicity, the output layer in this example is considered as a unitary weighted summation ($w_{1l}^3 = 1$ for $l = 1, 2, \cdots, S$) of the signals coming from the previous layer, just followed by a pure linear activation function. The total amount of coefficients to be tuned in backpropagation would be the sum of the number of weights ($F(N + 1 + S)$) and biases ($F + S$). For example, when considering four memory taps (5 neurons in the input layer), 8 neurons in the first hidden layer and 6 neurons in the second hidden layer, 88 weights and 14 biases totaling 102 parameters need to be tuned.

Generally speaking, and when comparing the ANN approach with the polynomial one (discussed in previous sections), the polynomials have inherent local approximating properties in contrast to the global approximation capability of ANNs, when modeling strongly nonlinear systems. In addition, when compared to classical models the ANN may adapt better to extrapolating beyond the zone exploited for parameter extraction [Gil07a].

There is not a universal recipe to set up the best ANN architecture, learning algorithm or activation function given a specific problem. Trial and error is frequently employed but some physical knowledge of the phenomena to be modeled can be important when optimizing the resources and aiming to reach the best modeling performance. However, some design considerations (architectures, activation functions, backpropagation detail and learning algorithms, metrics, etc.) found in literature and being experimentally validated and benchmarked will be given in Section 4.1.2 to assist in the modeling and compensation of RF transceiver impairments.

4.1.2 Design Considerations for Digital Linearization and RF Impairment Correction

The ANNs are considered as an alternative to complex Volterra-based nonlinear models that require an unaffordable complexity to characterize the RF impairments in highly demanding transceiver architectures such as massive MIMO. The FTDNN architecture, combined with a back propagation learning algorithm (BPLA), has been over the last ten years one of the most attractive approaches for dynamic nonlinear modeling. Another frequent type of FFNN being widely exploited to predict the behavior of the PA is the radial basis function neural network (RBFFNN) which can progressively keep increasing the number of neurons in the hidden layer until the desired performance is met as shown in [Isa05]. The RNNs, that have been preferred to model dynamic nonlinear systems with feedback paths or frequency-dependent phenomena, allow for better characterization of the interaction between input and output samples and the cross terms. However, these networks have traditionally employed lengthy training algorithms making difficult their implementation in real-time systems. While MLPNN variations are pre-

dominantly employed for PA modeling and DPD, the RNN are found in works dealing with I/Q modulator gain and phase imbalances or to apply CFR. When all the impairments, including strong PA nonlinearities, have to be solved with a common architecture, the RNN may be hardly implementable or will be outperformed by the MLPNN-based approaches as will be shown in the next subsections. The better the dynamic nonlinearities are modeled, the better the linearization performance will be.

ANN Architectures for Single-Antenna DPD

In order to extract amplitude and phase information from modulated complex waveforms, the ANNs need to consider operating with either complex-valued (CV) input signals, weights and activation outputs or real-valued (RV) double-inputs double-outputs (and real weights and activation outputs), i.e. in the form of multiple I and Q components. Complex-valued operation leads to heavy calculations and a longer training phase. In addition, the architectures that employ independent NN to separately model the AM/AM and AM/PM behavior may fail in the synchronous convergence of the two NN and thus tend to overtrain the fastest converging one [Raw10]. The RV FTDNNs, which combine I/Q RV processing with input time delay lines (TDLs) to handle memory effects (but not output-to-input TDLs as it would happen in a RNN), can offer better superior performance and easy baseband implementation when used for inverse modeling of PAs with strong nonlinearities and memory effects. As seen in [Raw10], these ANN utilize a similar structure to that shown in Fig. 4.3, but instead of a single input single output NN now we are having double-input double-output (I/Q inputs and outputs). In this case, the weighted summation at each first hidden layer neuron will include a sequence of input samples both for the I and Q components (this information will propagate throughout the NN according to the activation functions) and the output layer may have a non-unitary weighted summation at each I and Q output accounting for the contributions of each neuron output in the second hidden layer. This ANN has therefore a maximum of $2F(N+1)+SF+2S$ weights and $F+M+2$ biases (considering the notation in Fig. 4.3) that will be adjusted using feedforward backpropagation.

As previously introduced, the RNNs can be modified or unrolled in most of the cases in such a way so as to emulate a FFNN scheme where consolidated BPLAs are applied. A relevant design consideration is that choosing an ANN architecture without taking into account which are the sources that generate the nonlinearities may impact negatively on the performance or be highly resource inefficient. For instance, when modeling the PA nonlinear dynamic effects, it can be worth paying attention to the PA physical model to reflect output-to-input interactions or account for memory effects given a signal bandwidth. One example of the previous design considerations is found in [Mka11] and is displayed in Fig. 4.4.

In this example, nonlinear activation functions are used to model static PA nonlinearities (typically, the stronger the nonlinearities the more neurons that are required) while linear acti-

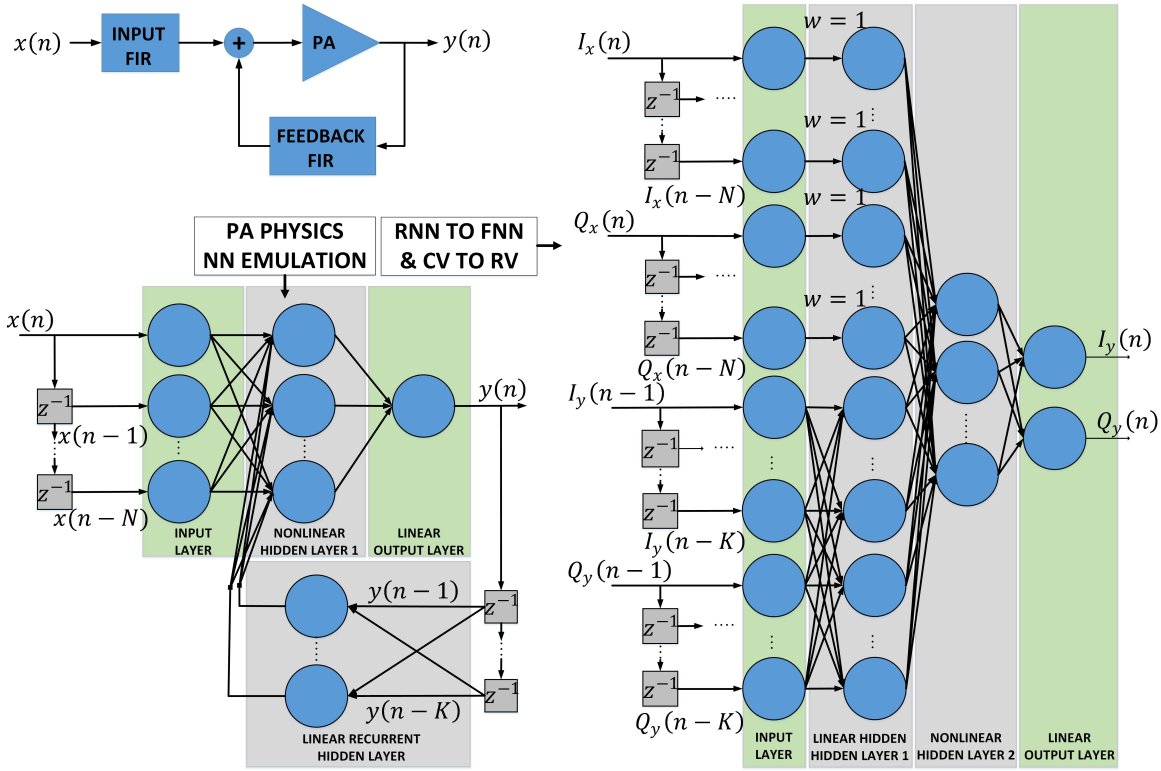


Figure 4.4: PA physics-aware CV RNN reformulated as a modified RV FTDNN.

vation functions will be used to model the feedback mechanism FIR filter (with memory depth K). The input FIR (input signal) models the memory effects which result from combining a wideband modulated signal with a non-flat response input matching network (N is the memory depth of the input signal). For FFN deployment, the connection between the initial input samples and the neurons is 1 (at first hidden layer). To train this NN, only measured past data from the PA output will be used at the input (not during validation since the delayed output data will be fed back once the necessary outputs are produced).

Moreover, some recent works have shown the benefits brought by adding envelope dependent terms as inputs to the ANN. [Jue17] propose a two-hidden layer RV FTDNN that includes one additional input based on the calculation of the modulus of the I and Q samples. This input is not only fed to the first hidden layer but directly to the output layer, all of which helps to improve the numerical stability and training convergence. [Wan19] inject additional envelope-dependent term combinations (i.e. between the modulus raised to the power of two, three, four and five) as inputs of a single hidden layer neural network (no direct connection between envelope related inputs and the output layer is enabled in this case). This obtains better modeling performance since these new terms are able of generating some desired even-order intermodulation terms throughout the neural network that cannot be obtained if only I and Q components are used as inputs.

ANN Training and Parameter Extraction Procedure

The ANN structure is typically trained with relevant I and Q baseband signals with expanded bandwidth and sample rate to allow for the DPD out-of-band compensation to fulfill ACPR requirements, and some additional envelope-dependent terms that can contribute to enhance the performance and the training speed and numerical stability. These ANN excitations lead to output signals which are typically compared with those taken from measurements at the PA output after RF-to-baseband downconversion, time alignment and gain compensation. In this batch-trained supervised learning environment, the BPLA is applied to tune the weights and biases given the selected parameter subset of layers, neurons and activation functions.

The previous works that have been highlighted to provide procedural indications mainly refer to the indirect learning (IL) architecture shown in Fig. 4.5 (left) that models the inverse response of the PA during the NN training and replicates the coefficients calculated in the forward path (NN inference). The better performing direct learning architecture (DL) that is shown in Fig. 4.5 (right), models the counteracting distortion signal to be added to the original input signal to compensate for the PA distortion.

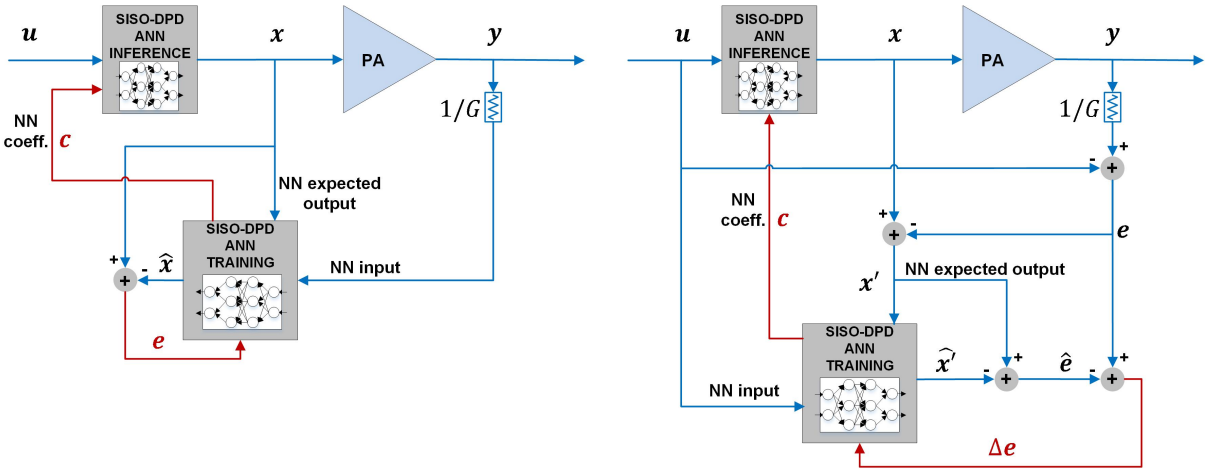


Figure 4.5: ANN-based DPD IL (left) and DL (right) architectures.

As introduced when comparing the CV- and RV-data operation in ANNs, using I and Q components features a significantly faster training by using real weights instead of using complex weights. The real-valued time delayed feedforward NN in Fig. 4.6 is now considered. According to the DL scheme, the input IQ pair is defined as the real and imaginary part of the NN input vector $u[n]$, and the output IQ pair will be an estimation of the real part and imaginary part of the NN expected output (target goal signal) $x'[n]$. The input IQ pair is further processed by a group of data functions (DF in Fig. 4.6) to provide richer basis to the ANN and improve the nonlinear modeling. The ANN DPD can benefit from the injection of envelope dependent terms (i.e., $|I_{in} + jQ_{in}|^k$ with $k \in \mathbb{N}$) [Wan19], angle dependent terms (i.e. $\operatorname{atan} 2\left(Q_{in}/I_{in}\right)^k$ with $k \in \mathbb{N}$),

or long-term memory effects modeling components like those described in [Teh12] (i.e., $s[n] = \frac{1}{K}\sum_{k=0}^{K-1}|I_{in}[n-k]+jQ_{in}[n-k]|^2$, where $s[n]$ is a sample-by-sample sliding integration window with K-sample length, where K determines the long-term memory depth). To enable dynamic nonlinear system identification, the LTI system built with tapped delay lines is added. In this architecture therefore, the ANN input signal u contributes to generate the $N = (2+V)(M+1)$ input dataset basis functions that result from applying the LTI system (with M memory depth) to the IQ input signals and to the V DF outputs with the exception of the long-term memory effect modeling DF (in such case, $N = (2+V-1)(M+1)+1$). The input dataset length relies on the length L of the IQ data being considered for modeling. The ANN features one input layer (with N inputs), two hidden layers with F and S neurons, and an output layer with 2 neurons. The neurons in the hidden layers employ the $\vartheta^1(\cdot)$ and $\vartheta^2(\cdot)$ nonlinear activation functions and a pure linear activation function is used for the output layer. This NN architecture has $NF+FS+2S$ weights and $F+S+2$ biases. In Fig. 4.6, $w_{i,j}^k$ is the j^{th} weight belonging to the i^{th} neuron in the k^{th} layer. Similarly, θ_i^k is the bias belonging to the i^{th} neuron in the k^{th} layer.

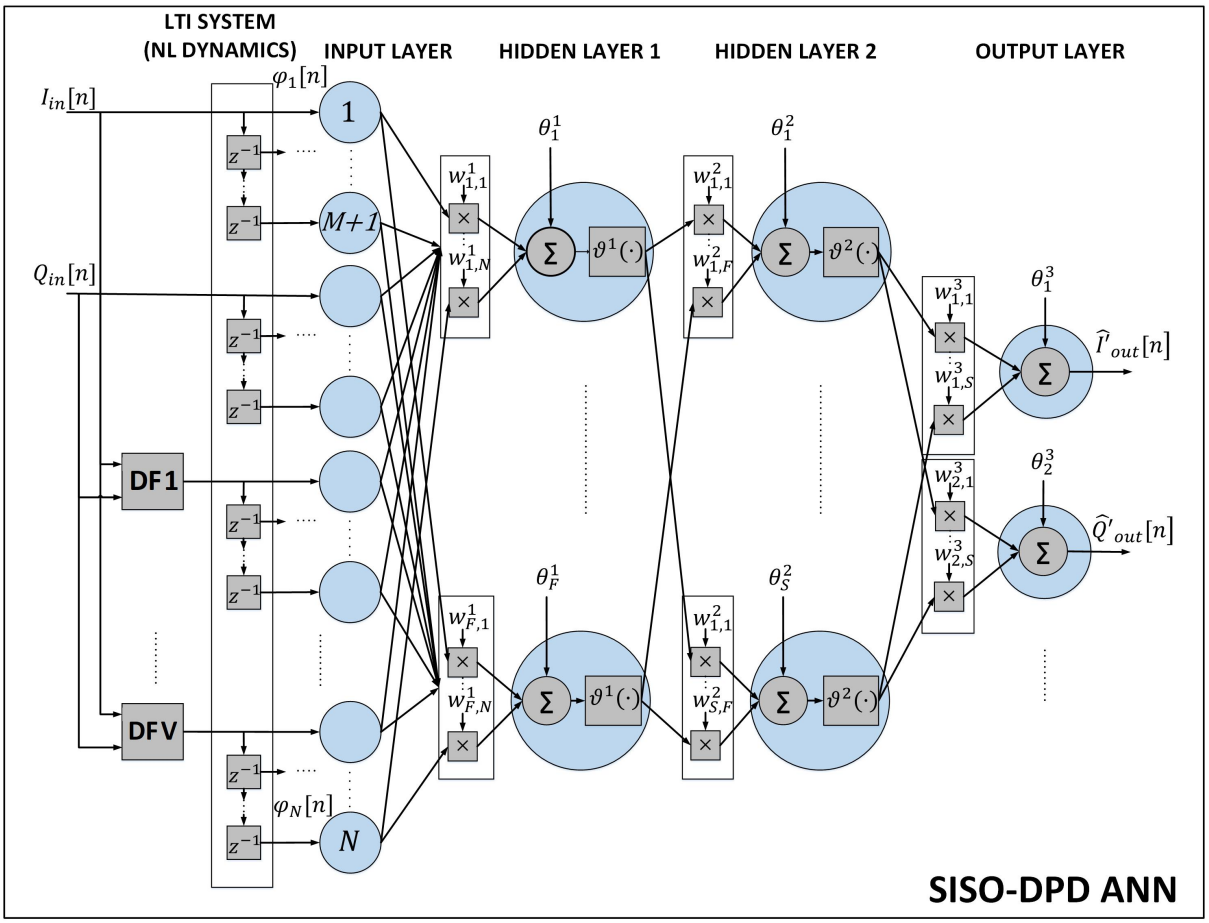


Figure 4.6: Two-hidden layer real-valued feed forward shallow ANN architecture.

$$\hat{I'}_{\text{out}}[n] = \sum_{k=1}^{S} w_{1,k}^{3}\ \vartheta^{2}\left(\sum_{j=1}^{F} w_{k,j}^{2}\vartheta^{1}\Big(\sum_{i=1}^{N} w_{j,i}^{1}\ \varphi_i[n] + \theta_j^{1}\Big) + \theta_k^{2}\right) + \theta_1^{3} \tag{4.3}$$

$$\hat{Q'}_{\text{out}}[n] = \sum_{k=1}^{S} w_{2,k}^{3}\ \vartheta^{2}\left(\sum_{j=1}^{F} w_{k,j}^{2}\vartheta^{1}\Big(\sum_{i=1}^{N} w_{j,i}^{1}\ \varphi_i[n] + \theta_j^{1}\Big) + \theta_k^{2}\right) + \theta_2^{3} \tag{4.4}$$

Selecting convenient initial weight and bias values for the ANN can be crucial to avoid training divergence or long training periods which do not learn significantly and thus deliver underperforming models. A general rule to be followed is to avoid extreme values (either the smallest or the largest) and symmetrical distribution of weights which make the neurons to perform similarly and thus provoke unnecessary redundancy and lower performance. If no initial knowledge is considered, the weights are chosen in a way that the input to the next activation function lies typically in the region between linear and saturated (see Fig. 4.2). Random initialization between -0.8 to +0.8 leads to a reasonably good starting point while the values below -1 and above +1 are avoided since the neuron learning will be very slow or will be stopped.

At every training epoch (or iteration) there are both a forward pass, where the error cost function is calculated with the outputs of the ANN and the desired outputs, and a backward pass that calculates the increment to be applied to the NN weights and biases to minimize that cost function. Having an adaptive digitally-assisted linearization or RF impairment compensation technique in the baseband modem puts some constraints on the number of epochs used for learning to reach the desired modeling performance. A categorization of the fast BPLA techniques is found in [Raw10]. In the first category, we can find the heuristic techniques (more detail is provided in [Hay09]) which are derivations from the analysis of the standard steepest descent algorithm. The gradient descent with momentum (GDM), which prevents from falling into bad local minima, can be included in this category together with variants such as the variable learning rate (GDA), the momentum and adaptive learning rule (GDX) and the resilient back-propagation (RP). The standard numerical optimization techniques are in the second category (the information is expanded in [Hag96]). This category includes; 1) the conjugate gradient-based techniques such as the Polak-Ribiere (CGP), the Fletcher-Powell (CGF), the Powell-Beale (CGB) and the scaled conjugate gradient (SGC), 2) quasi-Newton algorithms such as the Broyden-Fletcher-Goldfarb-Shanno (BFG) and the one step secant (OSS), and 3) the Levenberg-Marquardt (LM) algorithm. The LM combines the gradient descent and the Gauss-Newton methods, and is vastly used in ANN to minimize the cost function in DPD-related applications given its fast convergence properties which are paired with good modeling performance and fair implementation complexity [Hay09].

If the Levenberg-Marquardt (LM) BPLA is used in this ANN to calculate the NN coefficients, a cost function J, where

$$\begin{aligned} J &= \frac{1}{2L_b}\sum_{n=1}^{L_b}\left(I'_{\text{goal}}[n]-\hat{I'}_{\text{out}}[n]\right)^2+\left(Q'_{\text{goal}}[n]-\hat{Q'}_{\text{out}}[n]\right)^2 \\ &= \frac{1}{2L_b}\sum_{n=1}^{L_b}\left(e_I[n]-\hat{e}_I[n]\right)^2+\left(e_Q[n]-\hat{e}_Q[n]\right)^2 \\ &= \frac{1}{2L_b}\sum_{n=1}^{L_b}\left|\Delta e[n]\right|^2 \end{aligned} \tag{4.5}$$

is minimized for every training input data batch of length L_b samples. This forward-backward process is repeated until the desired modeling performance is achieved or the NN fails in generalization. When going backward, the LM modification to the Gauss-Newton method [Hag94] minimizes J with respect to a parameter $\boldsymbol{c}$ with the weights and biases, which is updated as

$$\boldsymbol{c}^{n+1} = \boldsymbol{c}^{n} - \left[\boldsymbol{J}^T\boldsymbol{J}+\beta\boldsymbol{I}\right]^{-1}\boldsymbol{J}^T\Delta\boldsymbol{e} \tag{4.6}$$

where

$$\boldsymbol{c} = \left[w^1_{1,1}\dots w^1_{F,N}\theta^1_1\dots\theta^1_F\dots\dots w^3_{1,1}\dots w^3_{2,S}\theta^3_1\theta^3_2\right] \tag{4.7}$$

where $\boldsymbol{I}$ is the identity matrix, β is a learning rate parameter and $\boldsymbol{J}$ is the Jacobian matrix calculated over the error vector e with respect to $\boldsymbol{c}$ as

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial\Delta e[1]}{\partial w^1_{11}} & \frac{\partial\Delta e[1]}{\partial w^1_{12}} & \dots & \frac{\partial\Delta e[1]}{\partial\theta^3_1} & \frac{\partial\Delta e[1]}{\partial\theta^3_2} \\ \frac{\partial\Delta e[2]}{\partial w^1_{11}} & \frac{\partial\Delta e[2]}{\partial w^1_{12}} & \dots & \frac{\partial\Delta e[2]}{\partial\theta^3_1} & \frac{\partial\Delta e[2]}{\partial\theta^3_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial\Delta e[L_b]}{\partial w^1_{11}} & \frac{\partial\Delta e[L_b]}{\partial w^1_{12}} & \dots & \frac{\partial\Delta e[L_b]}{\partial\theta^3_1} & \frac{\partial\Delta e[L_b]}{\partial\theta^3_2} \end{bmatrix} \tag{4.8}$$

whose elements can be computed as described in [Jar18a].

The learning rate parameter is introduced to guarantee the convergence of the BPLA. Both learning rate and momentum terms can be introduced in the algorithms minimizing the estimation error. As found in classical DPD learning schemes, the learning rate controls the convergence speed. If it is too small, the convergence is very slow and reaching the desired modeling performance requires more epochs, while if it is too high it can make the algorithm diverge. In [Ber96], the best learning rate is found from the Hessian matrix of the input signal that, however, changes significantly with time and is computationally complex to track. In [Mka11], the authors propose applying to the learning factor either an increasing or decreasing rate at every epoch depending on whether the error between the network outputs and the desired output is respectively meeting the desired performance or not. Given the fact that even with an appropriate choice of the learning rate the BPLA may suffer from convergence to a local optimum, in order to

better approach a global optimum the authors follow the procedure by [Pla86] and include the momentum term into to the BPLA. This factor adds relative contribution of the current and past errors to the current change of the estimated parameters in the shape of an oscillatory descent solution. Therefore, the ANN modeling performance can be benchmarked choosing first between a static or a dynamic learning rate and then between a static or a dynamic momentum term. For example, one could start employing a low value learning rate (e.g., one or two orders of magnitude below 1) and depending on whether the cost function or performance index J in Eq. 4.5 had increased or decreased, the learning rate could either be multiplied or divided, respectively, by an additional factor set to 10.

Procedures and Validation Methodologies

To achieve a better trade-off between the modelling performance and processing complexity, the following procedures should be followed in finding an ANN based solution while no universal rule exists now.

- *Input data memory depth*: The memory depth of the input signals is chosen typically by benchmarking different depth values in terms of modeling performance or NMSE (characterizing the error between the expected output and that obtained by the ANN) and complexity. For instance, a setting of memory taps which is 2 dB below the best NMSE attained could be the optimal one if the number of taps could be significantly reduced and the NMSE obtained was sufficient according to the application requirements. The knowledge of the PA physics and the designer expertise evaluating the PA response under a wideband modulated signal can be of help to reduce the number of cases to be evaluated.

- *Number of hidden layers and neurons*: The universal approximation theorem by [Cyb89] has been previously introduced in this chapter to justify the capacity of a single hidden layer ANN to approximate any nonlinear function with any desired error with a convenient activation function. However, the theorem does not specify the best solution in terms of learning time (or epochs), implementation complexity, number of hidden neurons or generalization capability with non-trained data, and assumes noise-free training data which is not always met in practice when the data is taken from measurements. Several works analyzing the performance of two-hidden layer versus single hidden layer ANN schemes concluded that the two-hidden layer ANNs provide better generalization and stability against the training data noise. [Che90] proved that adding the second hidden layer filters out the measurement noise that the single hidden layer operation does not (since it models the noise instead of filtering it out). [Hus93] proved that a two-hidden layer network may require a lower overall number of neurons than a single hidden layer scheme to approximate a modeling function. In general, it is hard to find generalized deterministic approaches to

choose the number of hidden layers since they would need to be validated under a massive amount of different data sets. Therefore, the final empirical selection of hidden layers may be driven in the end by trading-off the overall size of the ANN or the complexity, the learning time and the modeling accuracy [Tho16].

There is not a specific rule in selecting the optimal number of neurons at each hidden layers despite the fact that the stronger the PA nonlinearities, the higher the number of neurons in hidden layers (with nonlinear activation functions) will be. However, the complexity of the ANN can be set by evaluating the generalization error obtained when combining the bias-variance dilemma [Gem92] and the cross-validation technique [Sto78], [Hay09]. The bias error can be seen as how far from the expected data is the output data of the ANN model when using the training or estimation data set. A high bias error is indicative of underfitting. The variance error comes from the sensitivity to small variations over the training data set when the output of the ANN is evaluated with the validation data set. A high variance error is an indicator showing that the ANN is modeling the random noise in the training data instead of the intended outputs and thus it is overfitting. When the number of hidden neurons increases, the bias error typically decreases and the variance error increases. These parameters are taken into account in a back-propagation algorithm that learns in stages moving from the realization of simpler to more complex mapping functions as the training session progresses and the iteration or epoch number increases. By using this procedure, the training session is stopped periodically (i.e. every five epochs) and, given the obtained ANN weight and bias values, the model is tested on the validation subset at each of these periods. The MSE of the estimation during training decreases monotonically for an increasing number of epochs while in validation the MSE curve first decreases to a minimum and then increases (the learning algorithm starts modeling the noise given the training data set). An early MSE minimum could define the stopping point at which the ANN parameters are selected but, however, the number of epochs is typically increased beyond this stopping point to check whether the early MSE minimum is local or not and then choose the most convenient stopping point. A few improved versions of the cross-validation method and pruning procedures are found in [Hay09].

- *Activation functions*: Regarding the type of activation functions, there is neither a systematic approach to set the suitable function in the hidden layers. As shown in previous examples, for the output layer a pure linear function is typically used summing up the outputs of hidden neurons and linearly mapping them at the output. According to [Kar06], the linear activation functions are used typically for regression while nonlinear activation functions are used for input-output modeling. It is well-known that faster training can be achieved by using antisymmetrical activation functions such as the antisymmetric hyperbolic tangent. Again, the benchmarking of different activation functions for the hidden layers can help to determine which is the best option. Some information in this regard is

provided in the next subsection.

Some further quantitative analyses and evaluation results about the usability and performance of the ANN in behavioral modeling, DPD and RF impairment compensation applications are also given in [Gil20b]. Not only the impact of the ANN architecture parameter selection is assessed but also a comparison between ANN architectures with respect to classical polynomial-based approaches. The works in [Raw10] and [Mka11] are used to compare the RV FTDNN, RV RNN and polynomial-based architectures in terms of DPD modeling performance and stability, and compare the performance and training time with different BPLA algorithms, and also considering different learning and momentum rates, training signal data length or ANN activation functions. In terms of complexity of the network, these works also show that for multiple-layer ANNs the combination in number of neurons at the first hidden layer and the second hidden layer can be benchmarked in terms of NMSE (once the optimal number of neurons at the first hidden layer is set at a fixed number, the number of neurons in the second is then evaluated). The memory depth is also another parameter that is benchmarked. Different number of memory taps can be assessed where, typically, for RV inputs the same configuration will be employed while in RV RNN different taps configurations between input and feedback signals can be used depending on the physics to be modeled.

Finally, [Gil20b] also shows several ANN approaches that have been proposed over the last 15 years to overcome several limitations of the classical PAPR reduction schemes. An overview of ANN CFR applications constituting a more convenient alternative to signal scrambling-based probabilistic schemes requiring explicit side information such as SLM, PTS, tone injection or ACE, and to signal distortion techniques such as peak cancellation or clipping and filtering, is also reported including architectural detail and the advantages in terms of either PAPR reduction performance or implementation complexity (or both).

Neural Network Modeling Use Case

Evaluating ANN design considerations for GaN PA behavioral modeling and DPD

By having in mind the aforementioned procedures, validation methodologies and the benchmarks found in literature, it is easier to find the ANN structure and parameters that can lead to achieve better results than random selection of the parameters. Non-optimized parameters will lead to finding many constant response saturated nonlinear neurons that do not contribute to the modeling of the dynamic nonlinear behavior while the contrary happens in optimized designs. In this section, an experimental validation of the design principles is provided. With the test setup described in Section 3.2.3, multiple PA input-output measurements have been conducted to characterize the nonlinear behavior of the test setup GaN class-J PA operating at

875 MHz center frequency with high-PAPR (12-14 dB) FBMC waveforms both having 20 MHz and 80 MHz signal bandwidth.

Table 4.1 shows the number of coefficients and PA behavioral modeling performance benchmark when considering different parameter configurations of a RVFTDNN in terms of memory terms, number of neurons in single or dual hidden layer configuration, envelope dependent terms. The activation functions and different BPLAs have also been accounted. The measurements have been processed in the MATLAB environment where the RVFTDNN has been trained by setting up a maximum of 100 training epochs and where, after verifying results generalization with different waveforms, the worst case NMSE has been annotated.

Table 4.1: GaN PA Nonlinear Modeling With RVFTDNN.

MEM. TERMS	NEURONS IN HL	$\|I+jQ\|^T$ TERMS	WEIGHTS + BIASES	ACTIVATION FUNCTION IN HL[a]	BPLA	NMSE (dB)
2	10	-	72	tansig	LM	-25.35
4	10	-	112	”	”	-31.91
6	**10**	**-**	**152**	”	”	**-33.69**
10	10	-	232	”	”	-34.16
18	10	-	392	”	”	-35.07
6	4	-	62	tansig	LM	-26.39
6	8	-	122	”	”	-32.74
6	16	-	242	”	”	-35.52
6	28	-	422	”	”	-36.47
6	**10**	**T=1**	**212**	**tansig**	LM	**-37.48**
6	10	T=1,2,3	332	”	”	-37.66
6	10	T=1,2,3,4,5	452	”	”	-40.10
6	10	T=1,3,5	332	”	”	-36.77
6	10	T=2,4	272	”	”	-35.04
6	10	T=1	212	logsig	LM	-37.75
6	10	”	212	radbas	”	-36.90
6	10	”	212	satlins	”	-31.80
6	10	T=1	212	tansig	BFG[b]	-24.02
6	10	”	212	”	CGB[b]	-24.30
6	10	”	212	”	RP[b]	-23.73
3	**5-10 (2HL)**	**T=1**	**132**	**satlins+tansig**	**LM**	**-37.38**
3	5-10 (2HL)	T=1	132	satlins+tansig	LM[c]	-21.42
6	**12-24 (2HL)**	**”**	**590**	**satlins+tansig**	**LM[c]**	**-36.70**
20	20	T=1	1262	tansig	LM[c]	-39.43

[a] Activations functions that have been considered (in order of appearance in the table): Hyperbolic Tangent Sigmoid, Log-Sigmoid, Radial Basis, Symmetric Saturating Linear.

[b] Epoch limit increased to explore performance. BFG: Broyden, Fletcher, Goldfarb, and Shanno, CGB: Conjugated Gradient with Powell/Beale restarts, RP: resilient backpropagation.

[c] Results for 80 MHz bandwidth signal.

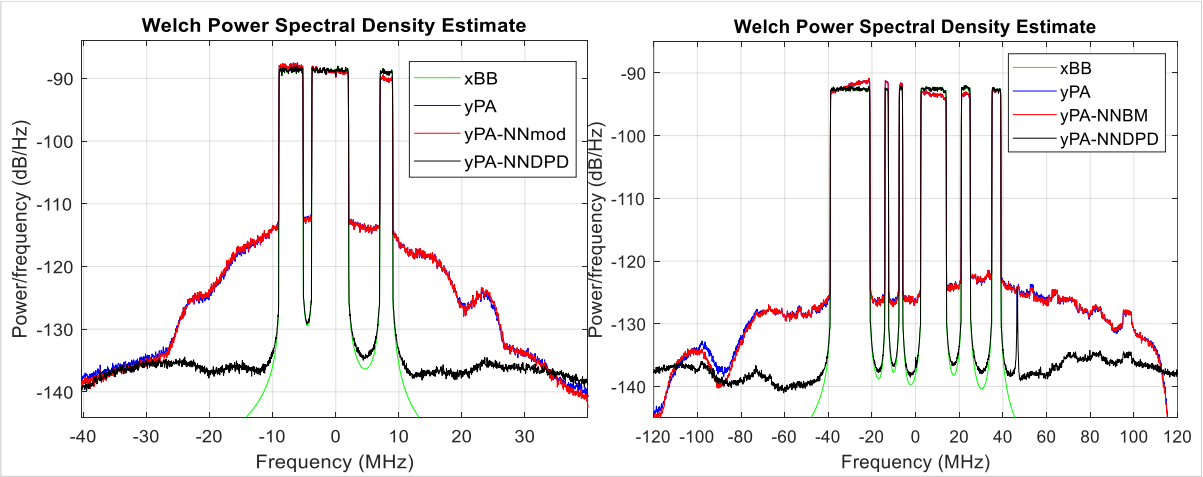


Figure 4.7: PA and DPD behavioral modeling for a GaN PA with strong nonlinearities and operated with 20 MHz (left) and 80 MHz (right) bandwidth signals.

The effectiveness of several design considerations that have been reported is demonstrated in Table 4.1. For instance, it highlights the importance of considering complex envelope dependent terms, the convenience of using two hidden layers instead of only one, the suitability of employing nonlinear antisymmetric activation functions or the performance of the LM algorithm when compared with other options. Fig. 4.7 also demonstrates the suitability of the best fitting configurations for 20 MHz and 80 MHz bandwidth signals.

When the DPD oversampled versions of the original 20 MHz and 80 MHz bandwidth signals are compared (with 153.6 MHz and 368.64 MHz DPD baseband sample rates after oversampling by 5 and 3, respectively), it can be seen that the DPD complexity (i.e., number of NN weights and biases) can increase significantly for the latter configuration when similar nonlinear modeling performance is targeted. On top of the added measurement noise due to increasing the signal bandwidth, the PA can also show stronger (long term) memory effects, some effects may be better defined or identified with the higher measurement acquisition rates, and the DPD receiver is also more susceptible to capture spurious signals falling within the DPD measurement bandwidth.

4.2 Dataset Reduction Techniques for SISO Neural Network Digital Predistorters

4.2.1 Introduction to ANN Input Dataset Reduction

Section 4.1 has shown how the artificial neural networks (ANN) are gaining momentum in the digital predistorters (DPD) thanks to their inherently good approximation capabilities. Under strong or complex power amplifier nonlinearities, however, both the size of the input dataset and the complexity of the ANN topology can increase and lead to long training periods which are unaffordable in fast-changing waveform scenarios like those proposed for 5G or 6G. In this section, combining training data length reduction mechanisms together with feature selection and feature extraction techniques is proposed to significantly reduce the ANN dataset size, the number of coefficients at the first hidden layer, and to significantly shorten the ANN DPD coefficients update time.

By considering the ANN presented in Section 4.1.2 and Fig. 4.6, we can now define the LxN input dataset matrix $\boldsymbol{U} = \big(\boldsymbol{\phi}[0], \cdots, \boldsymbol{\phi}[n], \cdots, \boldsymbol{\phi}[L-1]\big)^T$, where $\boldsymbol{\phi}^T[n] = \big(\varphi_1^{\boldsymbol{u}}[n], \cdots, \varphi_j^{\boldsymbol{u}}[n], \cdots, \varphi_{(2+V)(M+1)}^{\boldsymbol{u}}[n]\big)$ can be seen as a vector containing all the specific dataset features (or basis functions) $\varphi_j^{\boldsymbol{u}}[n]$ linked to the ANN input signal $\boldsymbol{u}$ in Figure 4.5-right after applying also the data functions and LTI system in Figure 4.6. The techniques that are presented aim at reducing the two $\boldsymbol{U}$ matrix dimensions. By reducing the number of basis or dataset features N, both the number of coefficients at the first hidden layer and the training time are reduced. The latter will be much further impacted when applying dataset length reduction techniques in advance as shown in Section 4.2.4.

4.2.2 Dataset Basis Reduction with Feature Selection Technique

The Orthogonal matching pursuit (OMP) technique has been presented in Section 3.1.3 as an *a priori* offline study to determine the best basis functions to linearize the PA. This process is applied once and the basis selection will be kept for a given scenario (i.e. fixed PA output power, waveform and bandwidth). Now we could consider applying OMP to the ANN input dataset $\boldsymbol{U}$, which has N features after considering the IQ signals and data function augmentations and applying the LTI system to account for memory effects. But instead of accounting for the PA output signal y, in the algorithm we could now use the expected ANN output signal x'. Seeking to maximize the selection modeling performance (or to maximize the reduction factor for a given modeling performance goal) another feature selection algorithm has been considered.

In [Bec18b], the authors proposed the doubly OMP (DOMP) scheme that adds a Gram-Schmidt orthogonalization step into the basis selection procedure which contributes to better NMSE performance. The procedure detailed in Algorithm 6 highlights the differences with respect to the OMP algorithm detailed in Section 3.1.3, and shows how to apply DOMP to reduce the ANN input dataset features or variables by a factor of RF_d. Unlike OMP, DOMP adds deflation to the chosen columns so when a basis function is chosen, those still to be selected are orthogonal with respect to the chosen ones. Gram-Schmidt is performed by first obtaining a projection vector $\boldsymbol{\rho}$ of the selected regressor into each one of previously orthogonalized regressors in $\boldsymbol{Z}$. This projection is then used to decorrelate the basis functions with the selected regressor as per step 14 in Algorithm 6, where $\otimes$ is the Kronecker product operator. The selection step used in DOMP is different from that used in OMP, since it accounts for the minimum residual error after orthogonalization.

Algorithm 6 Doubly Orthogonal Matching Pursuit

1: **procedure** DOMP ($\boldsymbol{x}'$,$\boldsymbol{U}$,RF_d)
2: initialization:
3: $\boldsymbol{e}^{(0)} = \boldsymbol{x}' - \widehat{\boldsymbol{x}'}^{(0)}$; with $\widehat{\boldsymbol{x}'}^{(0)} = 0$
4: $\boldsymbol{\Upsilon}^{(0)} \leftarrow \{\}$
5: $\boldsymbol{Z}^{(0)} \leftarrow \boldsymbol{U}_{L\times N}$
6: **for** $m = 1$ **to** N/RF_d **do**
7: $\boldsymbol{Z}_{\{i\}}^{(m-1)} \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\leftarrow} \frac{\boldsymbol{Z}_{\{i\}}^{(m-1)}}{\|\boldsymbol{Z}_{\{i\}}^{(m-1)}\|_2}$
8: $i^{(m)} \leftarrow \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmin}} \underset{\boldsymbol{w_i}}{\min}\|\boldsymbol{e}^{(m-1)} - \boldsymbol{Z}_{\{i\}}^{(m-1)}\boldsymbol{w_i}\|_2^2 \approx \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmax}}|\boldsymbol{Z}_{\{i\}}^{(m-1)^H}\boldsymbol{e}^{(m-1)}|$
9: $\boldsymbol{\Upsilon}^{(m)} \leftarrow \boldsymbol{\Upsilon}^{(m-1)} \cup i^{(m)}$
10: $\boldsymbol{w} \leftarrow \left(\boldsymbol{U}_{\{\boldsymbol{\Upsilon}^{(m)}\}}^H \boldsymbol{U}_{\{\boldsymbol{\Upsilon}^{(m)}\}}\right)^{-1}\boldsymbol{U}_{\boldsymbol{\Upsilon}^{(m)}\}}^H \boldsymbol{x}'$
11: $\widehat{\boldsymbol{x}'}^{(m)} \leftarrow \boldsymbol{U}_{\{\boldsymbol{\Upsilon}^{(m)}\}}\boldsymbol{w}$
12: $\boldsymbol{e}^{(m)} \leftarrow \boldsymbol{x}' - \widehat{\boldsymbol{x}'}^{(m)}$
13: $\boldsymbol{\rho}^{(m)} \leftarrow \boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)^H}\boldsymbol{Z}^{(m-1)}$
14: $\boldsymbol{Z}^{(m)} \leftarrow \boldsymbol{Z}^{(m-1)} - \boldsymbol{\rho}^{(m)} \otimes \boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)}$
15: **end for**
16: **return** $\boldsymbol{\Upsilon}$
17: **return** $\boldsymbol{U}_{\{\boldsymbol{\Upsilon}\}}$
18: **end procedure**

4.2.3 Dataset Length Reduction with Batch Selection Methods

Several sample selection methods (SSM) that reduce the computational complexity of DPDs have been introduced in Section 3.1.2. A new data batch selection (BS) method is now presented. The BS method takes the relevant waveform data batches in $\boldsymbol{u}$ depending on the characteristics of the PA output $\boldsymbol{y}$. Two variants of the BS method are considered: consecutive BS (CBS) and

sparse BS (SBS). The procedure has the following steps:

1. $\boldsymbol{u}$ and $\boldsymbol{y}$ are divided into a number of data batches N_b which is an integer multiple number of the waveform length reduction factor RF_l to be applied. The number of batches to be selected is $N_s = N_b/RF_l$.

2. With CBS, a sliding window of consecutive samples with length $L_s = L/RF_l$ is shifted over $\boldsymbol{u}$ and $\boldsymbol{y}$, with $L_b = L/N_b$ shifting step, to calculate both the NMSE, the ACEPR and the $\boldsymbol{u}$ signal mean power score metrics. When using SBS, a single batch sliding window moves with L_b shifting step and all the metrics are calculated individually for every consecutive batch (N_b in total). The BS algorithm is in charge of sorting the expanded batch indices (CBS) or the single batch indices (SBS) in order of importance according to featuring simultaneously the worst NMSE and ACEPR values, and the highest mean power to avoid the impact of the noise in the nonlinear modelling. The maximum score in a given metric will be equivalent to the overall number of either expanded (CBS) or single (SBS) score calculations.

3. With CBS there will be just a single batch selection according to the highest batch score metric while in SBS the indices are sorted descending according to their batch score and the selection is done with the highest scoring ones. After selection the batch indices are sorted ascending to facilitate having more consecutive batches (which favors better characterization of memory effects).

4. The last step is to conform the index vector $\boldsymbol{\beta}$ that will include the indices of all the samples contained at every batch selected. For illustration purposes the application of CBS and SBS to single channel DPD input-output data is exemplified in Fig. 4.8. The reduced NN input waveform will be thus populated with either a consecutive portion of $\boldsymbol{u}$ in CBS or a sparse number of portions in SBS, as shown in Fig. 4.9.

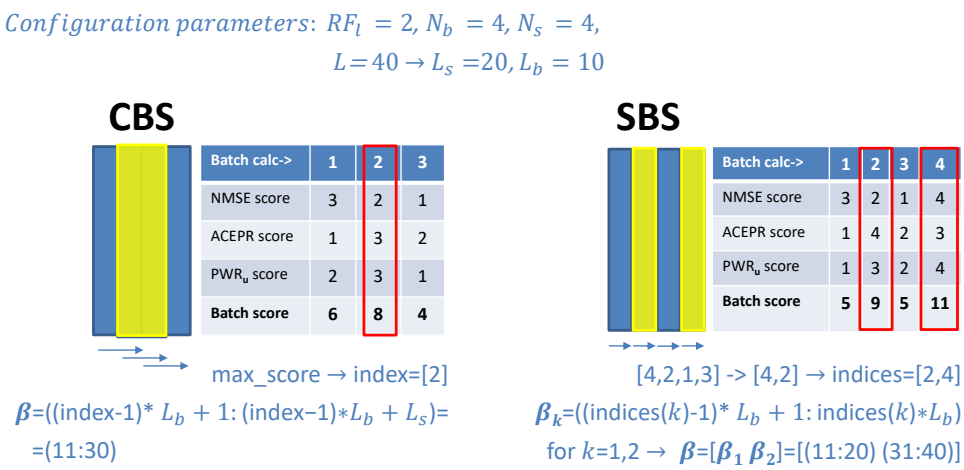


Figure 4.8: CBS and SBS batch selection procedure [LB21].

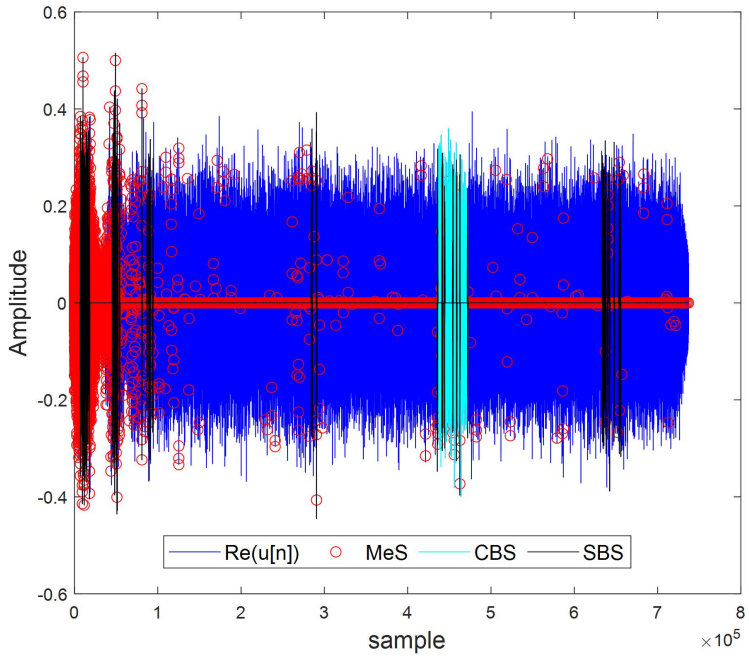


Figure 4.9: MeS, CBS and SBS input dataset length reduction methods (i.e. for $RF_l = 20$).

In the next section, the BS methods are compared to the evolved version of the histogram-based mesh selecting (MeS) method with multidimensional IQ mapping (i.e., to better account for wideband PA memory effects) presented in [Wan21a].

4.2.4 Experimental Results

The dataset reduction techniques have been benchmarked and validated with the MATLAB controlled hardware test bench shown in Fig. 4.10. A GaN HEMT class J PA which is operated at 875 MHz RF frequency and 28 dBm mean output power with 80 MHz bandwidth signals featuring 13-14 dB PAPR. The DPD neural network has been implemented in MATLAB. The dataset has 105 inputs/basis each with 737280 samples. These 105 inputs belong to the I and Q components and the augmented products until the sixth power, considering up to 12 consecutive delays, and a single long-term memory effects modeling component averaging the IQ input signal power over a finite window. The hidden layers have 20 neurons each and use hyperbolic tangent sigmoid activation functions.

The OMP and the length reduction techniques are evaluated independently in 4.11 and 4.12. The combined use is validated in Table 4.2. Both in 4.11 and Table 4.2 the reduction factors being considered are displayed after the dataset reduction technique acronyms. Fig. 4.11 shows how the complex ANN DPD performs better than the generalized memory polynomial (GMP) DPD when using the DPD coefficients that provide the maximum (MAX) performance. The ANN DOMP is better than the ANN OMP in 1-1.5 dB of NMSE but features similar ACPR, when

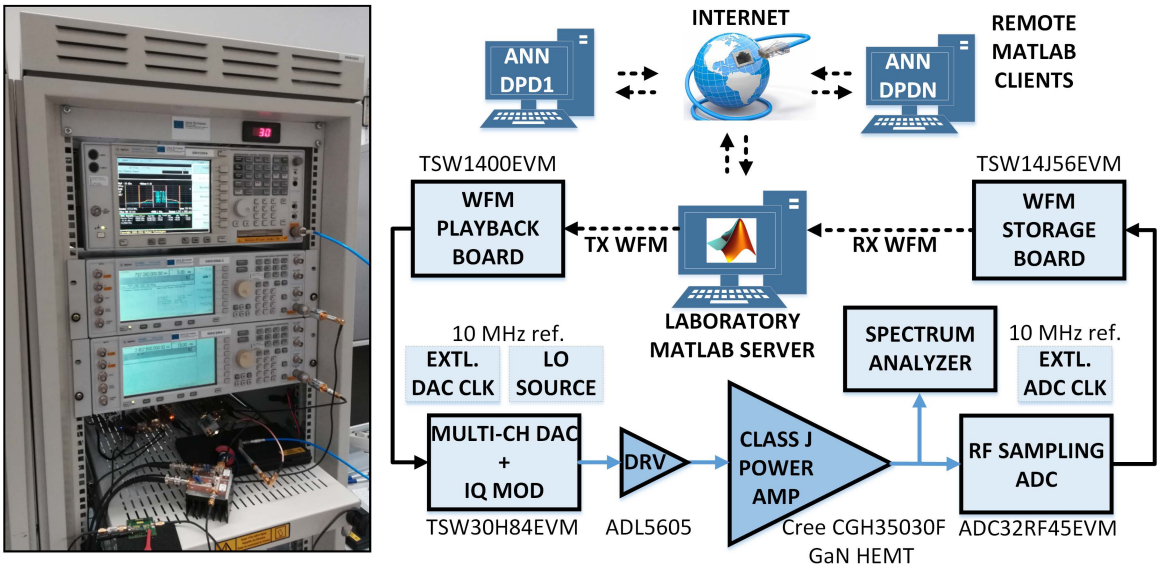


Figure 4.10: Remotely accessible digital linearization test setup.

using a reduction factor of 8. Using ANN DOMP with this reduction factor also outperforms the GMP DPD.

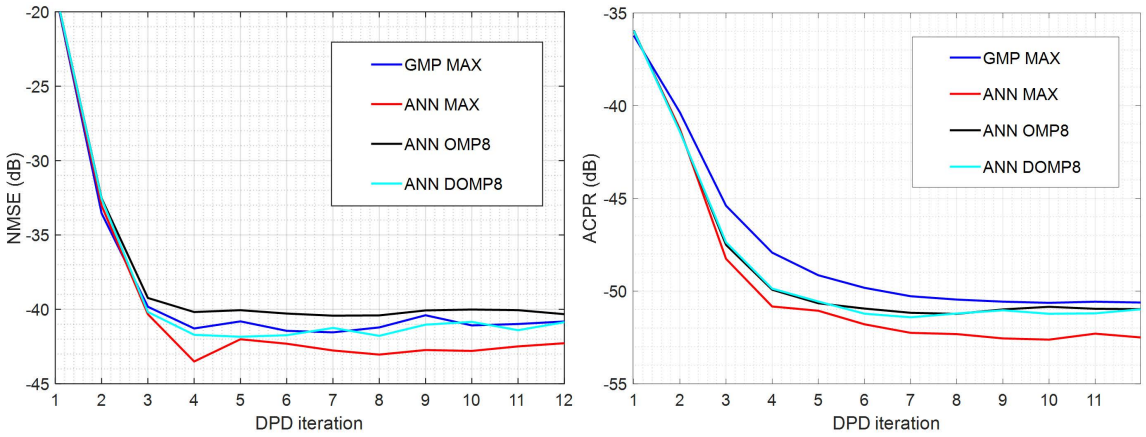


Figure 4.11: NMSE and ACPR vs DPD iteration under dataset basis reduction.

In 4.12, for lower reduction factors CBS is a good choice while in large reduction factors SBS is the best choice. The plot shows that even considering a reduction factor of 320 the NN still fulfills reasonable NMSE and ACPR. Finally, Table 4.2 shows how combining the two types of dataset reduction techniques can be beneficial. Combining basis reduction and length reduction in the NN implies an equivalent reduction both in terms of coefficients and training time (measured with MATLAB's tictoc and involving 8 DPD iterations), and better performance than only using the best BS dataset length reduction method (i.e. see SBS160 vs CBS40+DOMP4 in Table 4.2).

These results indicate that by using the input dataset reduction techniques the NN training

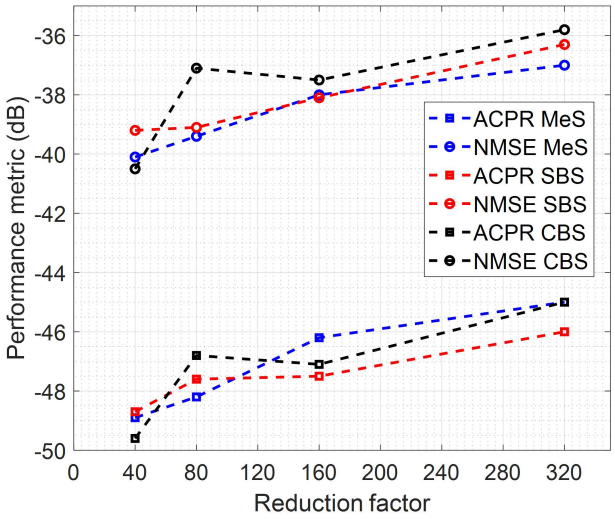


Figure 4.12: NMSE and ACPR vs dataset length reduction factor and method.

Table 4.2: Complexity, Timing and Performance Benchmark

Method	**Performance Benchmark**			
	DPD coeffs.	***8-Iteration DPD Update Time (s)***	***NMSE (dB)***	***ACPR (dBc)***
NN MAX	2582	8e4	-42.6	-52.5
NN DOMP4	1002	25e3	-41.4	-51.4
NN CBS40	2582	660	-40.5	-49.6
NN SBS160	2582	230	-38.1	-47.5
NN CBS40+DOMP4	1002	104	-39.2	-48.2
GMP MAX	322	320	-40.7	-50.6

time can be dramatically reduced and potentially be in the order of magnitude of the classical polynomial DPD adaptation time.

Chapter 5

Machine Learning Techniques for MIMO Digital Predistorters

5.1 Introduction: Overview and Challenges of ML application to MIMO DPD

In 5G and beyond radios, the increased bandwidth, the fast-changing waveform scenarios, and the operation of large array multiple-input multiple-output (MIMO) transmitter architectures have challenged both the polynomial and the artificial neural network (ANN) MIMO adaptive digital predistortion (DPD) schemes. This section proposes training data selection methods and dimensionality reduction techniques that can be combined to enable relevant reductions of the DPD training time and the implementation complexity for MIMO transmitter architectures.

The advent of 5G has brought deploying flexible waveform, numerology and frame design strategies together with increased bandwidth signals that operate in large-array transmitter architectures with RF impairments that cannot be handled properly by classical digital predistortion (DPD) linearizers [Fag19]. While in previous generation systems the DPD adaptation time was not a very limiting parameter, nowadays 5G multi-antenna transmitters may deliver fast-changing waveform configurations, either in terms of modulation, bandwidth occupation or power statistics, that require faster DPD coefficient adaptation. Such transmitters need also to handle complex linearization scenarios involving multiple PA input and output cross coupling effects and beam dependent PA loading effects (i.e., in high power circulator-less time division duplexing RF front ends), which aggravate the overall PA distortion [Hau17]. Fig. 5.1 shows the block diagram of one MIMO transmitter architecture for sub-7 GHz and depicts some undesired effects that may appear like cross couplings, I/Q gain and phase imbalances, DC offsets, in-band LO couplings and PA nonlinear distortion [Hän20b, Hän20a].

Several MIMO DPD behavioral models accounting for cross couplings have been presented over the last decade [Bas09, Saf11, Abd16], some of them capable to mitigate I/Q modulator

imbalances and dc offsets added to the crosstalk effects [Ant10, Ami14, Kha17]. The main concern of all the previous multivariate polynomial-based models is that a multi-input DPD is required in every transmit path and cannot be properly implemented for large-array MIMO transmitters. To overcome the requirement of including a multi-input DPD in every transmit path, a DPD system combining a single linear crosstalk and mismatch model block for the whole array and dual-input DPD models at each Tx path is proposed in [Hau18]. In addition, artificial neural networks (ANN) have been proposed in literature over the last years as an alternative to such higher complexity polynomial models for MIMO DPD [Zay10, Jar18a, Wan19]. Such schemes benefit from having only one ANN MIMO DPD block with as many inputs and outputs as transmitter baseband channels and PAs, which reduces the overall count of DPD coefficients for large arrays. The single DPD block approach is also adopted in [Yu19, Bri20b, Bri20c] for addressing the linearization of hybrid beamforming transmitter architectures at mm-waves.

The adaptive multi-antenna DPDs can become key building blocks in nowadays radio modems building massive digital MIMO or hybrid beamforming transmitter architectures, but only if designed to be computationally and power efficient. The increased signal bandwidth of 5G and B5G technologies (i.e., a few hundred MHz) together with the growth in computational complexity when accounting for multidimensional digital linearization, and the higher-speed DPD processing requirements of the new agile radio systems, challenges more than ever the implementation of multi-antenna DPDs in the digital front-end silicon devices. The combination of training data selection methods and behavioral modeling dimensionality reduction techniques

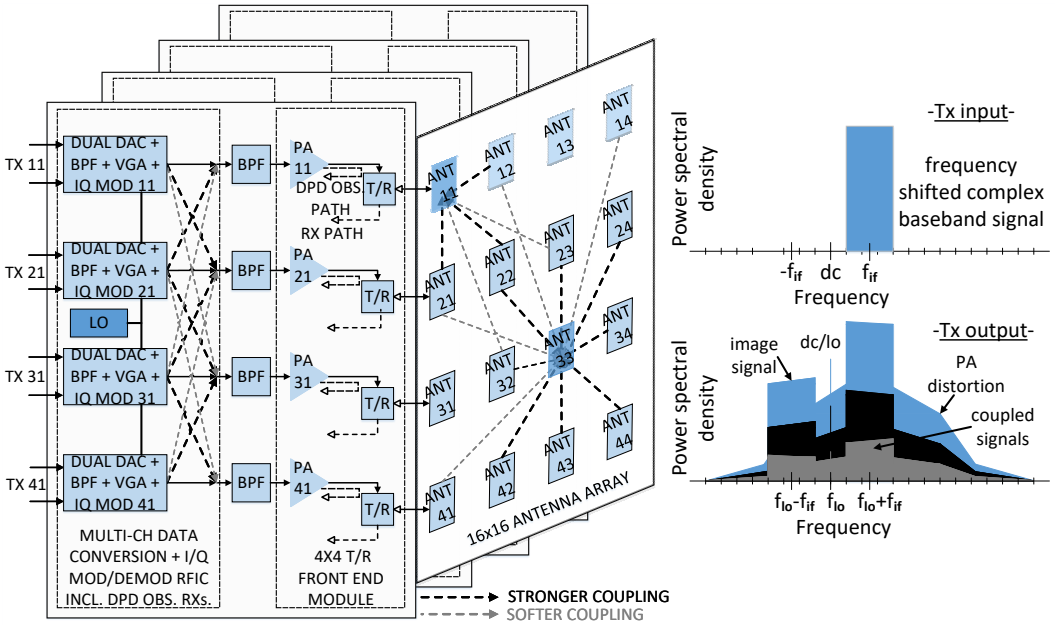


Figure 5.1: Multi-antenna circulator-less TX architecture (left) and illustration of the spectra of the complex baseband signal and the RF signal at a given PA output with typical RF impairments (right). Source: [Gil20b].

is a design strategy that may allow fast-adaptive well-conditioned MIMO DPD systems with optimized DPD processing complexity.

As seen in Section 3.1.2 the training data reduction methods reduce the memory requirements, the number of operations and the overall DPD adaptation time, by selecting convenient statistically representative training data samples. Several sample selection methods (SSM) have been reported in literature to reduce the computational complexity of polynomial- or piecewise-based DPD and to uncorrelate the observation errors in adjacent samples to get better performance [Wan17b, Zhu16, Kra20, Wan21a]. For ANN DPD, consecutive and sparse waveform batch selection mechanisms were proposed in [LB21] to reduce the training dataset length. It is worth noting that taking samples or batches of the waveform for DPD training cannot be done randomly when the waveform power statistics are unevenly distributed. These works target single-antenna transmitters but none is addressing the MIMO DPD training data selection considering the potential channel interplay. In MIMO DPD, only the dimensionality reduction (i.e., through basis selection) has been addressed.

Among the different dimensionality reduction techniques used for polynomial- or piecewise-based DPD, Chapter 3.1 has introduced some of the most commonly used feature selection techniques [Bar21] which are typically employed to reduce the number of DPD basis functions (and thus coefficients) in the real-time forward path. Instead, the feature extraction techniques are typically used in the observation path to ensure well-conditioned estimation and reduce the number of parameters required for DPD identification or adaptation [Pha19b]. In ANN-based, feature selection techniques have been used to reduce the input dataset and ANN complexity [Bar02], to tune the center of radial basis function neural networks [Che91] and to incrementally set the optimal number of hidden layer neurons [Hua12], but there are no works applying feature selection to reducing ANN DPD datasets.

Most of DPD complexity reduction works in literature apply to single channel transmit architectures and only a few combine training data selection and basis reduction techniques [LB21, Yan21]. Regarding multi-antenna DPD, the authors in [Zen14] use sparse estimation techniques to reduce the basis of MIMO Volterra-based polynomial models for moderate input-output crosstalk conditions. In [Bri20a], a piecewise closed-loop DPD including a pruning algorithm for faster adaptation is introduced, while in [Wan21c] singular vector decomposition (SVD) is applied for dimensionality reduction of multiuser MIMO arrays.

The combination of training data selection methods and dimensionality reduction techniques has not been holistically addressed in the literature for multi-antenna transmitter architectures with PA input and output cross couplings provoking strong nonlinear effects. Typically, such techniques have been evaluated in single-antenna systems or multi-antenna systems with negligible cross couplings that allow treating each PA independently. In this contribution, several training data selection methods and dimensionality reduction techniques are combined for use

in different polynomial and neural network direct learning MIMO DPD architectures to combat strong nonlinear effects and cross couplings. In previous works, the performance of the polynomial MIMO-DPD based on multiple input single output (MISO) polynomial DPD blocks is typically compared with that of the MIMO-DPD ANN based on a single MIMO ANN to predistort all the transmitter channels at once. The use of multiple MISO ANN blocks (imitating the MISO-DPD polynomial scheme with ANNs), whose use can be beneficial for a low number of antennas as demonstrated in this chapter, is also considered.

As proved in Section 5.6, the histogram-based sample or equation selection methods [Kra20, Wan21a] fail to deliver the expected performance in MIMO transmitter architectures when the channel of interest is contaminated by adjacent channels due to PA input and output cross talk. In this work, the waveform batch selection methods in [LB21], whose selection is based in batch performance metrics, is evaluated. More importantly, the more efficient uncorrelated equation selection (UES) method is proposed. Its novel application to data selection for least squares (LS) fitting, in this case applied to DPD, outperforms the previous techniques, supports operation under channel or antenna cross couplings, and runs faster.

Regarding the art in dimensionality reduction techniques, leveraging on the novel application of the orthogonal least squares (OLS) and PCA to MIMO DPD is proposed. The original OLS technique by Chen and Billings [Che89] is deployed for DPD basis selection. This technique has inspired similar approaches validated for single-antenna DPDs. The application procedure for use of the original OLS technique to reduce the polynomial model basis functions and ANN dataset features in MIMO DPD is detailed. OLS is able to outperform OMP in DPD basis selection and perform as well as other OMP later variants like the DOMP [Bec18b]. In polynomial-based DPD, PCA is commonly used in the identification subsystem to reduce the number of basis and avoid an ill-conditioned estimation while reducing the complexity of the LS calculation [Gil20b]. The PCA technique is proposed to reduce the number of features in large ANN MIMO DPD datasets. PCA is able to extract the hidden structure in high-dimensionality datasets, produce linear combinations of dataset features that help the ANN to reach the desired results for a given structure, and reduce the complexity both in training and inference thanks to reducing the number of weights needed in the ANN first hidden layer. As demonstrated in Section 5.6, PCA and OLS can be used to trade-off between ANN performance and reduced complexity (and faster adaptivity), respectively.

This chapter formulates and gives full visibility on how the methods for faster MIMO DPD adaptivity are applied to each of the three direct learning architectures considered, which is a contribution frequently obviated in the literature but necessary to allow reproducibility. The DPD adaptivity or some training aspects in ANN-based digital predistorters are commonly not addressed in the literature, and their limitations in comparison to polynomial DPDs are sometimes omitted. The techniques are combined and experimentally validated in a 2×2 MIMO

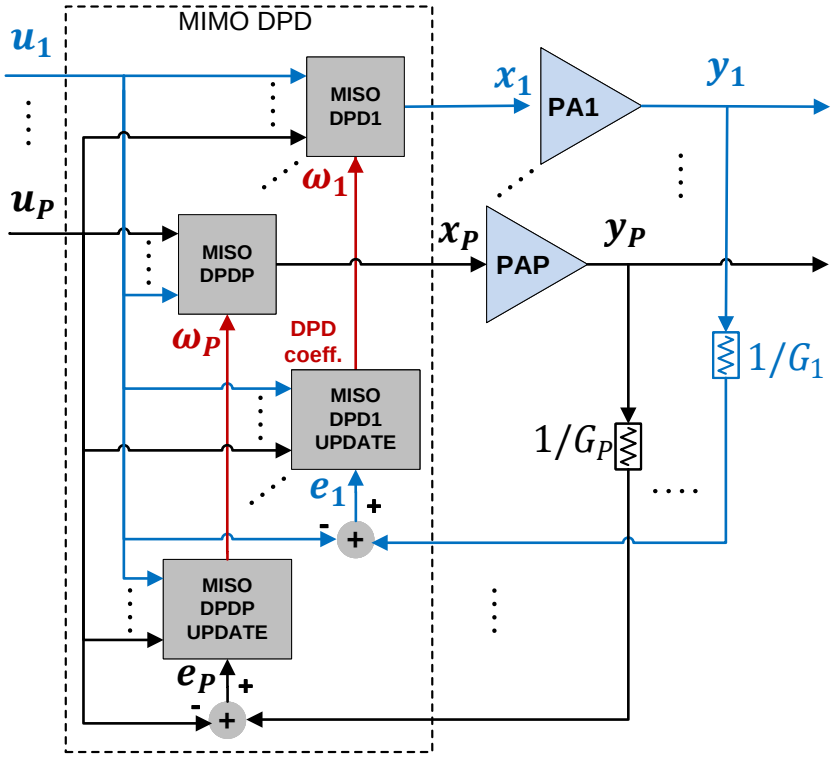


Figure 5.2: MIMO DPD direct learning architecture with independent polynomial MISO DPD blocks.

laboratory test bench for the three MIMO DPD schemes under the presence of strong cross couplings and nonlinear effects.

5.2 Direct Learning Multi-Antenna DPD Schemes

5.2.1 Polynomial-Based MIMO DPD Scheme

The block diagram of the polynomial MIMO DPD closed-loop adaptive architecture is shown in Fig. 5.2. The direct learning approach is again chosen since it is more robust against noisy PA output observations and can better avoid the offset of the coefficient vector from optimal values, when compared to the indirect learning approach which models the inverse PA response [Bra15]. In the forward path of this direct learning MIMO DPD scheme, we have one multiple input single output (MISO) DPD block per PA or antenna, whose input-output relationship can be described as

$$x_i[n] = u_i[n] - d_i[n] \tag{5.1}$$

where $x_i[n]$ is the predistorted signal at the output of the i^{th} MISO DPD block, with $i = 1, \cdots, P$, P is the number of PAs or antennas (or MIMO baseband signals in a digital MIMO transmitter),

$u_i[n]$ is the i^{th} MIMO baseband signal (it is noted that the MIMO input signals are uncorrelated), and $d_i[n]$ is the PA distortion signal to be modeled. The polynomial model will not only account for the i^{th} MIMO baseband signal but all the rest to deal with any PA input and output cross coupling and nonlinear effect. The expansion of the generalized memory polynomial (GMP) model for MIMO scenarios with nonlinear crosstalk (GMPNLC) in [Ami14] is used in this section to model the PA distortion. For the sake of simplicity, $d_1[n]$ is defined for a 2×2 MIMO scenario as

$$\begin{aligned} d_1[n] = &\sum_{o=1}^{O_1} \sum_{h=1}^{O_1-o+1} \sum_{m1=1}^{M1_1} \sum_{m2=1}^{M2_1} \alpha 1_{o,h,m1,m2}\ u_1[n-m_1] \\ &\big|u_1[n-m_1-m_2]\big|^{2(o-1)} \big|u_2[n-m_1-m_2]\big|^{2(h-1)} \\ &+ \sum_{o=1}^{O_1} \sum_{h=1}^{O_1-o+1} \sum_{m1=1}^{M1_1} \sum_{m2=1}^{M2_1} \alpha 2_{o,h,m1,m2}\ u_2[n-m_1] \\ &\big|u_2[n-m_1-m_2]\big|^{2(o-1)} \big|u_1[n-m_1-m_2]\big|^{2(h-1)}, \end{aligned} \tag{5.2}$$

where O_1 is a polynomial order, $M1_1$ and $M2_1$ are memory depths. The distortion signal for the second PA, $d_2[n]$, could consider different polynomial orders and memory depths O_2, $M1_2$, $M2_2$ if deemed necessary but it is equivalently defined (i.e., by swapping indices between the $u_i[n]$ input signals).

It is possible to rewrite (5.1) in matrix notation as

$$\boldsymbol{x_i} = \boldsymbol{u_i} - \boldsymbol{U_i}\boldsymbol{w_i}, \tag{5.3}$$

where $\boldsymbol{x_i} = \big(x_i[0], \cdots, x_i[n], \cdots, x_i[L-1]\big)^T$ and $\boldsymbol{u_i} = \big(u_i[0], \cdots, u_i[n], \cdots, u_i[L-1]\big)^T$, with $n = 0, \cdots, L-1$, are the MISO DPD block output predistorted signal and the i^{th} MIMO channel baseband signal input, respectively, and $\boldsymbol{U_i} = \Big(\boldsymbol{\phi_i}[0], \cdots, \boldsymbol{\phi_i}[n], \cdots, \boldsymbol{\phi_i}[L-1]\Big)^T$ is the $L \times N_i$ behavioral modeling data matrix, with L being the number of samples and N_i being the number of basis functions or the order of the i^{th} MISO DPD model, and where $\boldsymbol{\phi_i}^T[n] = \Big(\varphi_{i1}[n], \cdots, \varphi_{ij}[n], \cdots, \varphi_{iN_i}[n]\Big)$ is the vector containing the specific basis functions $\varphi_{ij}[n]$ with $j = 1, \cdots, N_i$ that will be applied to the input data to constitute $\boldsymbol{U_i}$. The vector of DPD coefficients with dimensions $N_i \times 1$ is $\boldsymbol{w_i} = (w_{i1}, \cdots, w_{ij}, \cdots, w_{iN_i})^T$. The mapping between the GMPNLC coefficients in (5.2) and the general DPD coefficients w_{ij} is direct for all the MIMO transmit channels. The total number of coefficients in the polynomial MIMO DPD will be $N_t = \sum_{i=1}^{P} N_i$.

In the MISO DPD update paths, the coefficients can be extracted iteratively finding the LS solution. At the k^{th} iteration the coefficients are obtained as

$$\boldsymbol{w_i}^{k+1} = \boldsymbol{w_i}^{k} + \mu\Big(\boldsymbol{U_i}^{\mathbf{H}}\boldsymbol{U_i}\Big)^{-1}\boldsymbol{U_i}^{\mathbf{H}}\boldsymbol{e_i} \tag{5.4}$$

where μ ($0 \leq \mu \leq 1$) is a weighting factor and $\boldsymbol{e_i} = \big(e_i[0], \cdots, e_i[n], \cdots, e_i[L-1]\big)^T$ is the $L \times 1$ DPD error vector defined as

$$\boldsymbol{e_i} = \frac{\boldsymbol{y_i}}{G_i} - \boldsymbol{u_i} \tag{5.5}$$

where G_i is the desired linear gain of the i^{th} PA, $\boldsymbol{y_i}$ is the signal at its output, and $\boldsymbol{u_i}$ is the i^{th} MIMO baseband signal. Considering the 2×2 GMPNLC model in (5.2), and assuming the same structure at each MISO DPD block ($O_1 = O_2 = O, M1_1 = M1_2 = M1, M2_1 = M2_2 = M2$), the order of the behavioral model after removing redundant products will be $N = P(P+1)(M1+1)(M2+1)$ and the overall number of the 2×2 MIMO DPD coefficients considering $\boldsymbol{w_1}$ and $\boldsymbol{w_2}$ will be $2N$.

Despite the number of necessary basis functions to model the distortion at every PA and account for cross couplings and other inter-branch nonlinear effects could be different (i.e., due to PA operation tolerances and non-symmetrical cross-effects between channels), the MIMO DPD models typically consider an initial common selection of sufficiently relevant basis functions at each MISO DPD block and therefore such blocks feature the same number of coefficients. The dimensionality reduction techniques presented in this work can contribute to reduce the final number of basis and preserve the relevant ones at every individual predistorter. Another aspect to be considered is that typically the MISO DPD models are defined having into account the input of every transmit channel. While this approach can be applied to small arrays, it is extremely inefficient for massive arrays given the fact that there will be elements that will be physically isolated and with no cross-effect interaction. Depending on the MIMO digital and analog front-end hardware, the number of different MIMO baseband signals to be routed into each MISO DPD block can be evaluated and optimized in factory for a given multi-antenna transceiver architecture layout.

5.2.2 Artificial Neural Network MIMO DPD Schemes

As shown in Fig. 5.3, the ANN-based MIMO DPD can either be built with P independent MISO DPD blocks, each with P inputs and a single output and producing a coefficient vector $\boldsymbol{c_i}$ (as in the polynomial approach), or with a more resource-efficient single MIMO DPD processing block with P inputs and P outputs and a unique coefficients vector $\boldsymbol{c}$.

In the MIMO DPD built with independent ANN MISO DPD blocks, every block is fed with all the baseband channel inputs $\boldsymbol{u_i}$ (in practice, only those with cross-effect interaction should be accounted) and every set of coefficients $\boldsymbol{c_i}$ is calculated to produce the ANN expected output (target signal) $\boldsymbol{x'_i} = \boldsymbol{u_i} - \boldsymbol{e_i}$ that precompensates for the distortion added by the corresponding PA and accounts for linear and nonlinear cross couplings from other MIMO channels. Therefore, following the notation in Fig. 5.3-top, the P sets of $\boldsymbol{c_i}$ coefficients are calculated independently to minimize every error increment $\boldsymbol{\Delta e_i} = \boldsymbol{e_i} - \boldsymbol{\hat{e}_i}$, where $\boldsymbol{e_i}$ is the residual linearization error

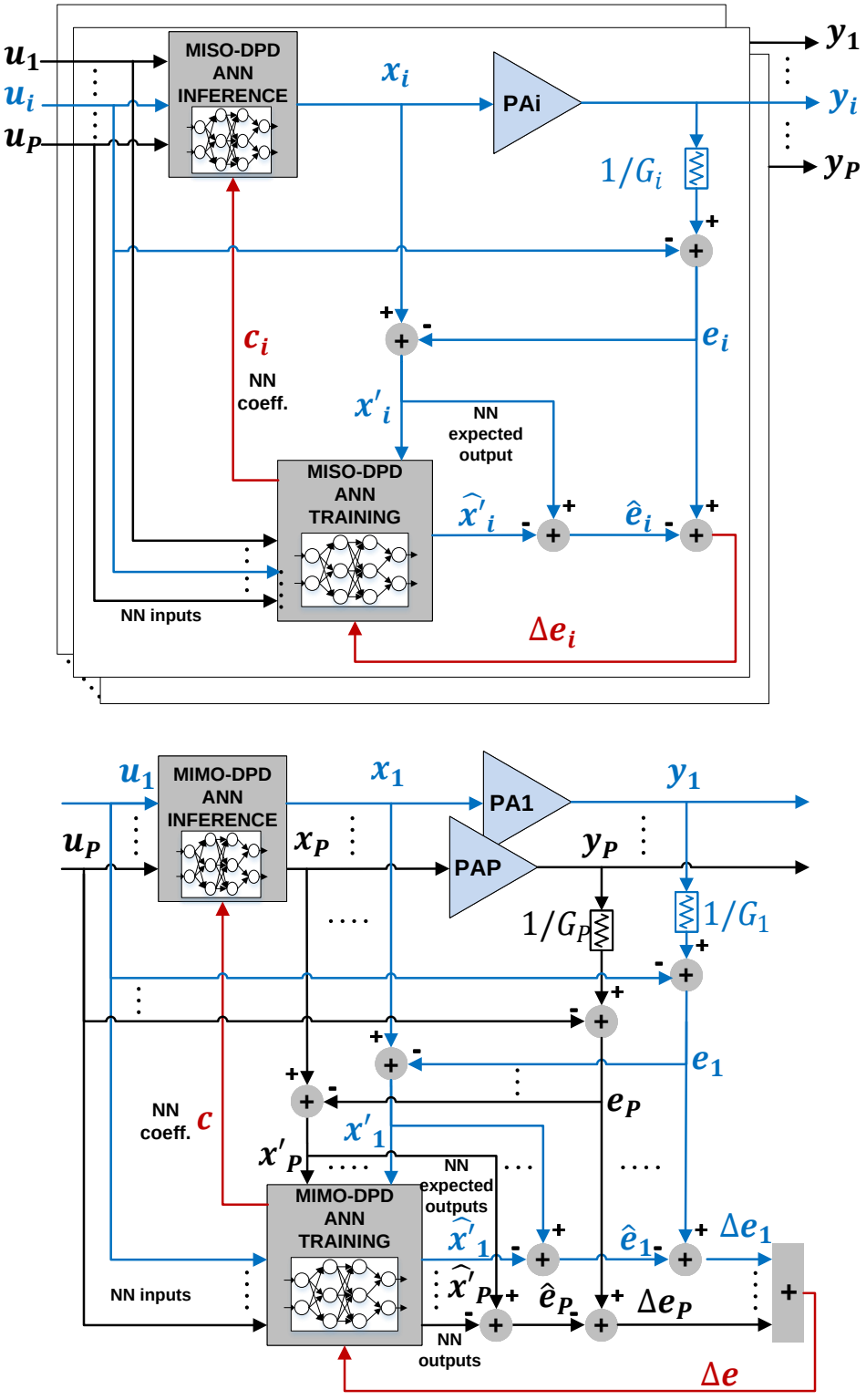


Figure 5.3: MIMO DPD direct learning architecture with independent ANN MISO DPD blocks (top) and with a single ANN MIMO DPD block (bottom).

vector at the i^{th} PA output as defined in (5.5) and $\hat{\boldsymbol{e}}_i = \boldsymbol{x}'_i - \hat{\boldsymbol{x}}'_i$ is the ANN estimated residual linearization error vector between the target signal and the estimated signal $\hat{\boldsymbol{x}}'_i$.

The MIMO DPD built with the single ANN MIMO DPD block in Fig. 5.3-bottom jointly generates all the precompensation signals minimizing the error between every $(\boldsymbol{e_i}, \hat{\boldsymbol{e}}_{\boldsymbol{i}})$ pair, and thus $\boldsymbol{c}$ is now calculated to minimize the sum of all the error increments $\boldsymbol{\Delta e_i}$. The topology shown in Fig. 5.4 is proposed for this architecture. It is an augmented version of the feedforward fully connected MIMO-RVTDNN in [Jar18a]. This ANN has real-valued input IQ pairs defined as the real and imaginary part of the input $u_i[n]$, where $I_{\text{in},i}[n] = Re\{u_i[n]\}$ and $Q_{\text{in},i}[n] = Im\{u_i[n]\}$, with $i = 1, \cdots, P$ (one input IQ pair per antenna), and real-valued output IQ pairs with $\hat{I'}_{\text{out},i}[n] = Re\{\hat{x'}_i[n]\}$ and $\hat{Q'}_{\text{out},i}[n] = Im\{\hat{x'}[n]\}$, where $\hat{x'}[n]$ is the ANN estimation of the targeted output $x'[n]$ with $I'_{\text{goal},i}[n] = Re\{x'_i[n]\}$ and $Q'_{\text{goal},i}[n] = Im\{x'_i[n]\}$ components. Every input IQ pair is augmented by several data functions (DF in Fig. 5.4) that enrich the ANN basis functions to improve nonlinear modeling. One of the augmentations that favor nonlinear modeling is using DFs that generate envelope dependent terms (i.e., $\left|I_{\text{in},i} + jQ_{\text{in},i}\right|^k$ with $k \in \mathbb{N}$) [Wan19]. Other DFs to be potentially considered provide powers of the IQ data (i.e., $\left|I_{\text{in},i}\right|^k$ and $\left|Q_{\text{in},i}\right|^k$) or angle dependent terms (i.e., $\left(\arctan(Q_{\text{in},i}/I_{\text{in},i})\right)^k$). It is well noted that such augmented products help the ANN to elevate the reachable modeling nonlinear order without having to increase the number of hidden layers using nonlinear activation functions. For proper dynamic nonlinear system identification, tapped-delay lines are added and applied to the input IQ pairs and to the DF outputs (M is the memory depth), delivering the N ANN inputs shown in Fig. 5.4. Long-term memory effect modeling components (i.e., $s[n] = \frac{1}{K}\sum_{k=0}^{K-1} |I_{ini}[n-k] + jQ_{ini}[n-k]|^2$) can also be introduced by the data functions but these inputs do not need to be replicated and delayed by the tapped-delay system.

All the previous ANN inputs constitute the input dataset features. Considering the input IQ data length L, the input dataset will have $L \times N$ size. The size of the dataset will be impacted by the complexity of the dynamic nonlinear effects to be modeled. As observed in Fig. 5.4, the input dataset $\boldsymbol{U} = (\boldsymbol{D_1}, \cdots, \boldsymbol{D_i}, \cdots, \boldsymbol{D_P})$ is built with the contributions from every input IQ pair or MIMO baseband channel routed into the MIMO DPD ANN. Considering P antennas, V data functions and M sample memory depth, after augmentation and delaying, every MIMO channel contributes with $(2+V)(M+1)$ dataset variables, totaling $N = P(2+V)(M+1)$ ANN inputs. We can now define $\boldsymbol{D_i} = \left(\boldsymbol{\phi_i}[0], \cdots, \boldsymbol{\phi_i}[n], \cdots, \boldsymbol{\phi_i}[L-1]\right)^T$, where $\boldsymbol{\phi_i}^T[n] = \left(\varphi_1^{\boldsymbol{u_i}}[n], \cdots, \varphi_j^{\boldsymbol{u_i}}[n], \cdots, \varphi_{(2+V)(M+1)}^{\boldsymbol{u_i}}[n]\right)$ can be seen as a vector containing all the specific dataset features (or basis functions) $\varphi_j^{\boldsymbol{u_i}}[n]$ linked to a given MIMO input signal $\boldsymbol{u_i}$, for $i = 1, \cdots, P$ and $j = 1, \cdots, (2+V)(M+1)$.

The proposed MIMO DPD ANN is therefore created with one input layer with N inputs, a first and a second hidden layer with F and S neurons, respectively, and an output layer with R neurons and outputs, with $R = 2P$. In the first and second hidden layer we have the $\vartheta^1(\cdot)$ and $\vartheta^2(\cdot)$ activation functions which are typically nonlinear functions such as the hyperbolic tangent sigmoid (well suited for nonlinear modeling), while a pure linear activation function is used at

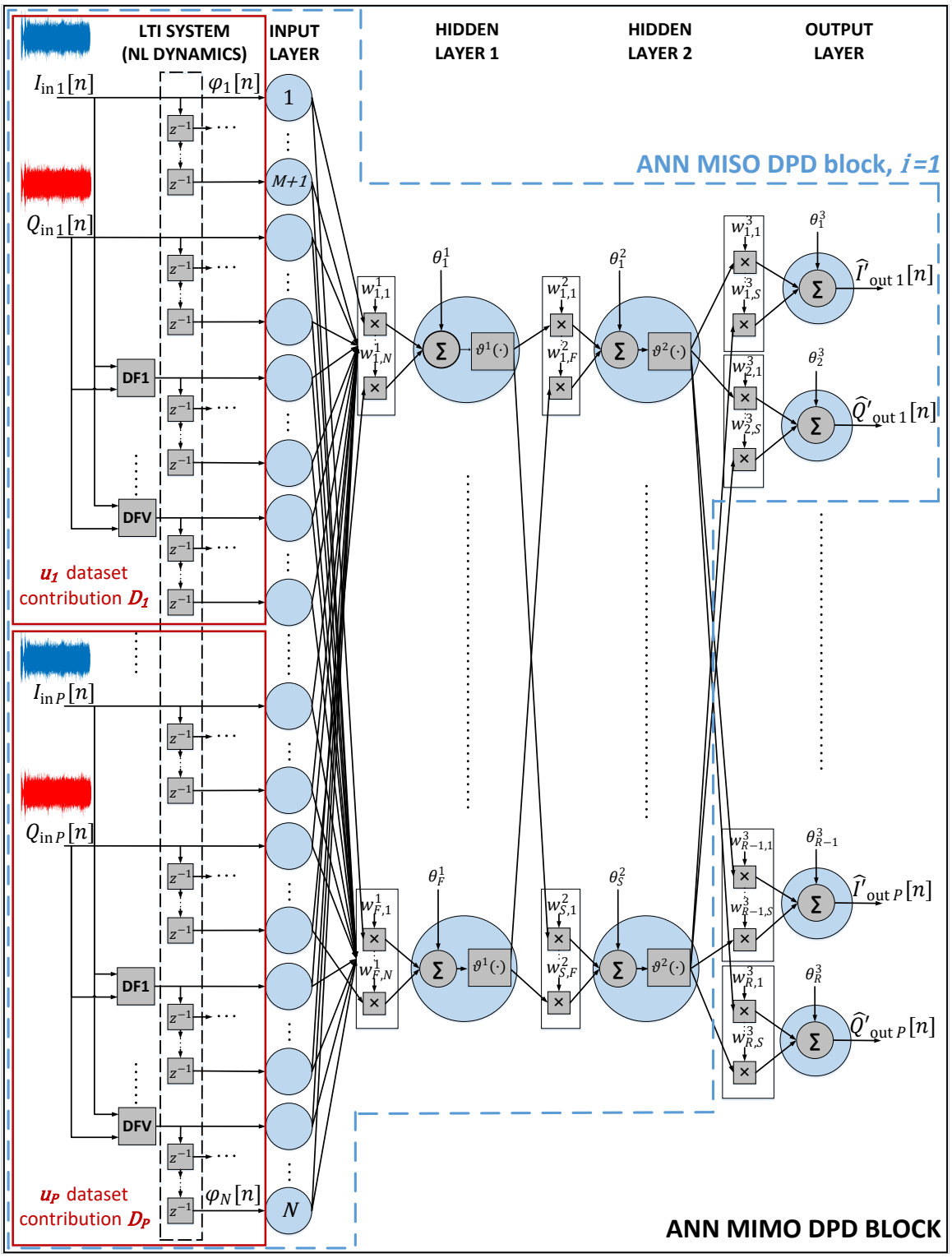


Figure 5.4: Two-hidden layer data function augmented fully connected RVTDNN for MIMO DPD application.

the output layer. The total number of ANN DPD coefficients in $\boldsymbol{c}$ is divided into $FN+SF+RS$ weights and $F+S+R$ biases.

In the approach with independent ANN MISO DPD blocks, the sum of all the $\boldsymbol{c_i}$ sets totals $P(FN+SF+2S)$ coefficients. While the complexity in terms of coefficients will be higher and the training of all the independent ANNs can take longer without hardware (HW) parallelization, the calculation of coefficients at each MISO DPD block will run faster and the performance will be also better due to dividing the problem into smaller ones (i.e., the coefficients at every independent ANN are tuned to minimize a single channel increment error vector and not the sum of potentially unbalanced different channel increment error vectors). In the notation used in Fig. 5.4, $w_{i,j}^{k}$ is the j^{th} weight belonging to the i^{th} neuron in the k^{th} layer. Similarly, θ_{i}^{k} is the bias belonging to the i^{th} neuron in the k^{th} layer. The MIMO DPD ANN output IQ pairs $\hat{I'}_{\text{out},i}[n]$ and $\hat{Q'}_{\text{out},i}[n]$ are calculated during training in the forward pass as

$$\hat{I'}_{\text{out},i}[n] = \sum_{k=1}^{S} w_{2i-1,k}^{3}\ \vartheta^{2}\left(\sum_{j=1}^{F} w_{k,j}^{2}\vartheta^{1}\Big(\sum_{l=1}^{N} w_{j,l}^{1}\ \varphi_{l}[n]+\theta_{j}^{1}\Big)+\theta_{k}^{2}\right)+\theta_{2i-1}^{3} \tag{5.6}$$

$$\hat{Q'}_{\text{out},i}[n] = \sum_{k=1}^{S} w_{2i,k}^{3}\ \vartheta^{2}\left(\sum_{j=1}^{F} w_{k,j}^{2}\vartheta^{1}\Big(\sum_{l=1}^{N} w_{j,l}^{1}\ \varphi_{l}[n]+\theta_{j}^{1}\Big)+\theta_{k}^{2}\right)+\theta_{2i}^{3}. \tag{5.7}$$

The MIMO DPD ANN output IQ pairs $\hat{I'}_{\text{out},i}[n]$ and $\hat{Q'}_{\text{out},i}[n]$ are calculated during training in the forward pass. The MIMO-RVTDNN coefficients $\boldsymbol{c}$ (ANN weights and biases) are then calculated with a backpropagation algorithm. Taking as reference the Levenberg-Marquardt (LM) [Hag94] algorithm and the fully integrated MIMO DPD ANN, the coefficients are calculated by minimizing the mean square error (MSE) cost function J in (5.8) for each training data batch of length L_b samples.

$$\begin{aligned} J &= \frac{1}{2L_b}\sum_{i=1}^{P}\sum_{n=1}^{L_b}\big(I'_{\text{goal},i}[n]-\hat{I'}_{\text{out},i}[n]\big)^{2}+\big(Q'_{\text{goal},i}[n]-\hat{Q'}_{\text{out},i}[n]\big)^{2} \\ &= \frac{1}{2L_b}\sum_{i=1}^{P}\sum_{n=1}^{L_b}\big(e_{Ii}[n]-\hat{e}_{Ii}[n]\big)^{2}+\big(e_{Qi}[n]-\hat{e}_{Qi}[n]\big)^{2} \\ &= \frac{1}{2L_b}\sum_{i=1}^{P}\sum_{n=1}^{L_b}\big|\Delta e_{i}[n]\big|^{2} = \frac{1}{2L_b}\sum_{n=1}^{L_b}\big|\Delta e[n]\big|^{2} \end{aligned} \tag{5.8}$$

It is also noted that, as shown in Fig. 5.3-bottom, $\boldsymbol{\Delta e_i} = \boldsymbol{e_i}-\boldsymbol{\hat{e}_i}$, where $\boldsymbol{e_i}$ is the residual linearization error vector at the i^{th} PA output as defined in (5.5) and $\boldsymbol{\hat{e}_i} = \boldsymbol{x'_i}-\boldsymbol{\hat{x}'_i}$ is the ANN estimated residual linearization error. When going backward, the cost function J is minimized with respect to the vector of coefficients $\boldsymbol{c} = \big(w_{1,1}^{1},\cdots,w_{F,N}^{1},\theta_{1}^{1},\cdots,\theta_{F}^{1},\cdots,w_{1,1}^{3},\cdots,w_{R,S}^{3},\theta_{1}^{3},\cdots,\theta_{R}^{3}\big)$ that

contains all the weights and biases of the MIMO RVTDNN. In contrast to the approach followed to calculate the coefficients in (5.4), to solve now a nonlinear least squares problem the LM algorithm updates $\boldsymbol{c}$ at every epoch k as

$$\boldsymbol{c}^{k+1} = \boldsymbol{c}^{k} - \left[\boldsymbol{J}^T\boldsymbol{J} + \beta\boldsymbol{I}\right]^{-1}\boldsymbol{J}^T\Delta\boldsymbol{e} \tag{5.9}$$

where $\boldsymbol{I}$ is the identity matrix, β is a damping factor, and $\boldsymbol{J}$ is the Jacobian matrix calculated over the increment error vector $\Delta\boldsymbol{e}$ with respect to $\boldsymbol{c}$ as

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial\Delta e[1]}{\partial w_{11}^1} & \frac{\partial\Delta e[1]}{\partial w_{12}^1} & \cdots & \frac{\partial\Delta e[1]}{\partial\theta_{R-1}^3} & \frac{\partial\Delta e[1]}{\partial\theta_{R}^3} \\ \frac{\partial\Delta e[2]}{\partial w_{11}^1} & \frac{\partial\Delta e[2]}{\partial w_{12}^1} & \cdots & \frac{\partial\Delta e[2]}{\partial\theta_{R-1}^3} & \frac{\partial\Delta e[2]}{\partial\theta_{R}^3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial\Delta e[L_b]}{\partial w_{11}^1} & \frac{\partial\Delta e[L_b]}{\partial w_{12}^1} & \cdots & \frac{\partial\Delta e[L_b]}{\partial\theta_{R-1}^3} & \frac{\partial\Delta e[L_b]}{\partial\theta_{R}^3} \end{bmatrix} \tag{5.10}$$

The forward-backward procedure is iterated until the required linearization performance is achieved or the ANN fails in generalization. This procedure is equivalently applied P times for the MISO ANN DPD. The $\boldsymbol{U}$ input dataset will be shared between all these independent ANNs but now the single target goals $\boldsymbol{x_i'}$ will be different according to the baseband channel of interest and the effects to be counteracted at each of the corresponding transmit chains. The cost function will not minimize the sum of error increments but the error increment $\boldsymbol{\Delta e_i}$ for a given channel instead (see Fig. 5.3-top).

5.3 Training Data Length Reduction

5.3.1 Batch and Equation Selection Methods

In Section 4.2 two batch selection methods named consecutive batch selection (CBS) and sparse batch selection (SBS) were presented. These methods were used to reduce the length of the ANN input dataset for nonlinear DPD modeling. These methods were compared with the SSM-inspired multidimensional IQ memory mesh selecting (MeS) method found in [Wan21a] that can be applied to the equations (or rows) of either polynomial behavioral modeling matrices or ANN input datasets. Compared to MeS, CBS and SBS could also achieve high reduction factors, similar NMSE performance but better ACPR thanks to employing batch selection score metrics linked to the out-of-band performance such as the ACEPR. One advantage of the batch selection methods is the versatility to adapt to different scenarios and properly balance the in-band and out-of-band distortion according to the requirements.

When these batch and equation selection techniques are applied to ANN MIMO DPD the following modifications can be applied to guarantee suitable time coexistence of the selected training data samples between every MIMO channel dataset contribution $\boldsymbol{D_i}$ in $\boldsymbol{U}$:

- For CBS and SBS, when the scores are calculated for every $(\boldsymbol{u_i}, \boldsymbol{y_i})$ pair, these are then summed, and a unique global batch index selection vector $\boldsymbol{\beta}$ is obtained for all the MIMO channel input dataset contributions. Given a length reduction factor RF_l, $\boldsymbol{\beta}$ will store the relevant L/RF_l training data indices that will be applied to the input-output data. This strategy can also be applied to the polynomial MIMO DPD case allowing for higher reduction factors. Under dominant strong cross couplings, only the mean power of the $\boldsymbol{u_i}$ signal batches is needed to select the right indices, which reduces the number of operations.
- With MeS the desired reduction factor RF_l is expanded by 50% and more equation indices than those initially required are selected. Once these indices are selected for all channels independently, only those appearing at all channels are taken and stored in the vector $\boldsymbol{\Theta}$. Finally, the number of coincident indices is reduced to provide the desired L/RF_l equation selection indices. This strategy is not required for the polynomial MIMO DPD.

These MIMO adaptations applied to the CBS, SBS and MeS techniques are also deployed in Section 5.5. From now onward, it will be assumed that the same length reduction factor is applied to any MISO polynomial or MISO ANN block (i.e., $RF_{li} = RF_l$, for $i = 1, ..., P$). As shown in the experimental results in Section 5.6, the traditional histogram-based methods can be underperforming for MIMO DPD application since the baseband sample selection made according to the histogram distribution of the baseband signal is no longer representative of what happens at PA level due to having uncorrelated couplings at the MIMO transmitter both before and after the PA. SBS and CBS, however, can be more suitable to select appropriate waveform segments thanks to tuning the performance scoring method to reflect, for instance, the batches where the waveforms are more impacted by the crosstalk effects.

5.3.2 Uncorrelated Equations Selection Method

A simple and computationally efficient equation selection method is proposed in this work for application to both the $\boldsymbol{U_i}$ polynomial behavioral modeling matrices and the $\boldsymbol{U}$ ANN input dataset. The procedure is detailed in Algorithm 7. It starts storing the first equation (i.e., matrix row) in $\boldsymbol{U}$ as the original master equation $\boldsymbol{r}$, and index one is stored in the equation selection index vector $\boldsymbol{\Theta}$. The scalar product between the master equation and the forthcoming equations is then applied. When one of such products is below the correlation factor (CF) threshold, the new master equation is stored together with its index. The process is repeated until all the equations are evaluated, the equation selection index vector $\boldsymbol{\Theta}$ is completed and it is then applied to $\boldsymbol{U}$ to reduce the number of equation or rows. The CF value to be set depends on the desired length reduction factor RF_l, i.e., the higher RF_l is, the smaller CF will be.

This method shows a good trade-off between implementation complexity and reduction performance. Such a procedure not only reduces the number of correlated equations, and thus

Algorithm 7 Uncorrelated equations selection (UES) method

procedure UES $(\boldsymbol{U}, CF)$
  initialization:
  $\boldsymbol{r} \leftarrow \boldsymbol{U}(1,:)$
  $\boldsymbol{\Theta}^{(0)} \leftarrow \{1\}$
  for $m = 2$ **to** L **do**
    if $|\boldsymbol{U}(m,:)\ \boldsymbol{r}^H| < CF$ **then**
      $\boldsymbol{r} \leftarrow \boldsymbol{U}(m,:)$
      $i^{(m)} \leftarrow m$
      $\boldsymbol{\Theta}^{(m)} \leftarrow \boldsymbol{\Theta}^{(m-1)} \cup i^{(m)}$
    end if
  end for
  return $\boldsymbol{\Theta}$
  return $\boldsymbol{U}_{\{\boldsymbol{\Theta}\}}$
end procedure

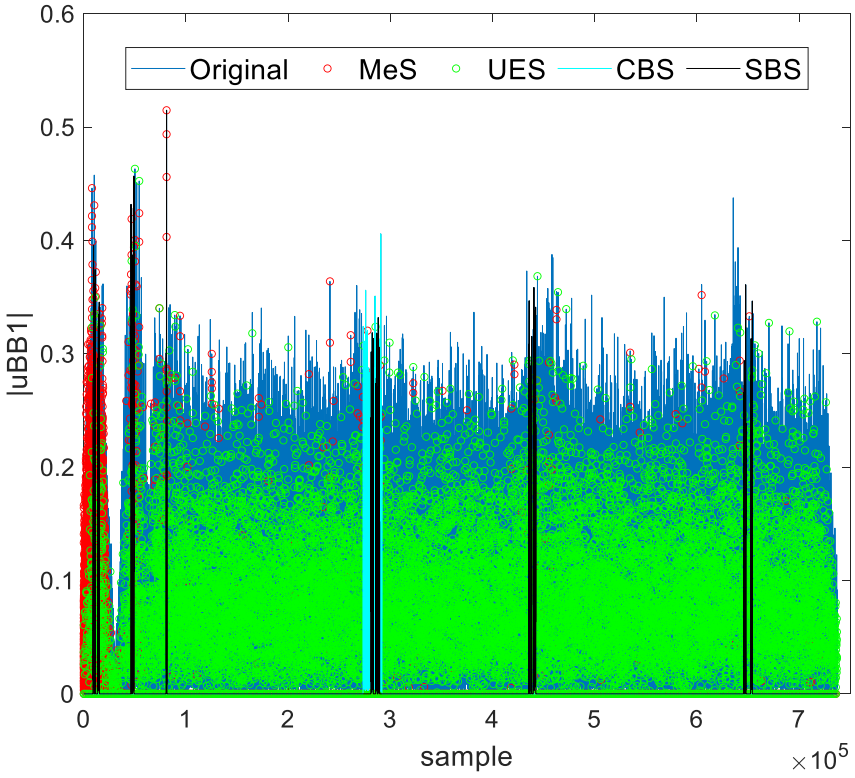


Figure 5.5: Training data length reduction methods application.

provides better system conditioning but it is also able to identify samples that are impacted by cross couplings. This is because the correlation between adjacent OFDM waveform samples for a given MIMO channel would be higher without crosstalk effects, and thus by sequentially selecting uncorrelated equations those highly impacted by crosstalk effects are stored. Fig. 5.5 shows an example on how the equation and batch selection methods look like when applied to

one input baseband signal with $RF_l = 40$. While MeS fills in most of the histogram bins with the samples or equations appearing initially in the training data, UES takes them in a more uniform fashion.

5.4 DPD Model Order Reduction

5.4.1 Orthogonal Least Squares

Several feature selection techniques have been proposed for DPD linearization to keep the most representative basis functions (or regressors) and thus reduce the modeling dimensionality [Bar21]. From the matching pursuit family, the orthogonal matching pursuit (OMP) [RT15] and the doubly OMP (DOMP) [Bec18b] have proved to be suitable approaches to trade-off linearity performance and model order reduction. In this work, the original orthogonal least squares (OLS) technique shown in Algorithm 8 was considered because it outperforms OMP and a challenging nonlinear scenario is being faced.

Algorithm 8 is defined generically. The cumulative modeling error $\boldsymbol{e}$ relies on a target signal $\boldsymbol{t}$. This target signal will be the PA output $\boldsymbol{y_i}$ when using the polynomial MIMO DPD scheme in Fig. 5.2 and $\boldsymbol{x'_i}$ when using ANN-based schemes like those in Fig. 5.3. As in OMP, in OLS it is possible to define a support vector ($\boldsymbol{\Upsilon}$) containing the indices of the best basis functions of a generic polynomial behavioral modeling matrix or ANN input dataset $\boldsymbol{U}$. At every iteration of the Algorithm 8 search, $\boldsymbol{\Upsilon}$ will be fed with the indices corresponding to the basis functions that better contribute to minimize the residual modeling error, which are sorted according to their relevance. For a given dimensionality reduction factor RF_d, the algorithm iterates N/RF_d times to fill $\boldsymbol{\Upsilon}$ with the required number of indices. This support set will be finally applied to the $\boldsymbol{U}$ matrix to obtain a reduced version having only the selected basis. Unlike OMP, OLS adds deflation to the chosen columns so when a basis function is chosen, those still to be selected are orthogonal with respect to the chosen ones. Gram-Schmidt is performed by first obtaining a projection vector $\boldsymbol{\rho}$ of the selected regressor into each one of previously orthogonalized regressors in $\boldsymbol{Z}$ (step 11). This projection is then used to decorrelate the basis functions with the selected regressor as per step 12 in Algorithm 8, where $\otimes$ is the Kronecker product operator. The selection step used in OLS is different from that used in OMP, since it accounts for the minimum residual error after orthogonalization [Blu07]. When comparing the OLS and DOMP algorithms, in order to update the residual error $\boldsymbol{e}^{(m)}$ in OLS, it is not necessary to calculate the values of the $\boldsymbol{w}$ coefficients as DOMP does (i.e., lines 10 to 12 in Algorithm 6 versus line 10 in Algorithm 8).

The specific usage of OLS both for the polynomial and the ANN-based MIMO DPD is detailed as follows. For polynomial MIMO DPD schemes, OLS will be applied at every MISO DPD block as $\boldsymbol{U_i^{OLS}} = \text{OLS}(\boldsymbol{y_i}, \boldsymbol{U_i}, RF_d)$ with $i = 1, \cdots, P$ (being P the total number of

Algorithm 8 Orthogonal Least Squares

procedure OLS $(\boldsymbol{t}, \boldsymbol{U}, RF_d)$
  initialization:
    $\boldsymbol{e}^{(0)} \leftarrow \boldsymbol{t}_{L\times 1}$
    $\boldsymbol{\Upsilon}^{(0)} \leftarrow \{\}$
    $\boldsymbol{Z}^{(0)} \leftarrow \boldsymbol{U}_{L\times N}$
  for $m = 1$ **to** N/RF_d **do**
    $\boldsymbol{Z}_{\{i\}}^{(m-1)} \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\longleftarrow} \frac{\boldsymbol{Z}_{\{i\}}^{(m-1)}}{\|\boldsymbol{Z}_{\{i\}}^{(m-1)}\|_2}$
    $i^{(m)} \leftarrow \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmin}} \min_{\boldsymbol{w_i}} \|\boldsymbol{e}^{(m-1)} - \boldsymbol{Z}_{\{i\}}^{(m-1)}\boldsymbol{w_i}\|_2^2 \approx \underset{i\notin\boldsymbol{\Upsilon}^{(m-1)}}{\operatorname{argmax}} |\boldsymbol{Z}_{\{i\}}^{(m-1)^H}\boldsymbol{e}^{(m-1)}|$
    $\boldsymbol{\Upsilon}^{(m)} \leftarrow \boldsymbol{\Upsilon}^{(m-1)} \cup i^{(m)}$
    $\boldsymbol{e}^{(m)} \leftarrow \boldsymbol{e}^{(m-1)} - (\boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)^H}\boldsymbol{e}^{(m-1)})\boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)}$
    $\boldsymbol{\rho}^{(m)} \leftarrow \boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)^H}\boldsymbol{Z}^{(m-1)}$
    $\boldsymbol{Z}^{(m)} \leftarrow \boldsymbol{Z}^{(m-1)} - \boldsymbol{\rho}^{(m)} \otimes \boldsymbol{Z}_{\{i^{(m)}\}}^{(m-1)}$
  end for
  return $\boldsymbol{\Upsilon}$
  return $\boldsymbol{U}_{\{\boldsymbol{\Upsilon}\}}$
end procedure

PAs, antennas or MIMO channels). For the ANN approach built with independent MISO DPD blocks, OLS is applied at every MISO DPD block as $\boldsymbol{U_i^{OLS}} = \text{OLS}(\boldsymbol{x_i'}, \boldsymbol{U}, RF_d)$, where $\boldsymbol{U} = (\boldsymbol{D_1}, \cdots, \boldsymbol{D_i}, \cdots, \boldsymbol{D_P})$ is the common MISO ANN's input dataset (i.e., for a small number of antennas all the MIMO baseband channels are routed into any MISO DPD ANN and therefore there is a single $\boldsymbol{U}$ dataset shared between all the MISO DPD ANNs before applying reduction. It is noted that every MISO DPD ANN would have different input datasets otherwise, depending on the physical coupling between channels at every PA or antenna. Note also that, for the sake of simplicity, it has been assumed that the same dimensionality reduction factor is applied to any polynomial or ANN MISO DPD block, i.e., $RF_{di} = RF_d$ for $i = 1, ..., P$. In the single-ANN MIMO DPD scheme, OLS is applied independently to every MIMO baseband channel input dataset contribution as $\boldsymbol{U^{OLS}} = \big(OLS(\boldsymbol{x_1'}, \boldsymbol{D_1}, RF_d), \cdots, OLS(\boldsymbol{x_P'}, \boldsymbol{D_P}, RF_d)\big)$, since we have a single dataset $\boldsymbol{U}$ but we need to handle simultaneously the P different target signals $\boldsymbol{x_i'}$. Finally, it is also important to remark that OLS can also be used for basis function or feature preselection to speed-up the UES method, as shown in Section 5.5, thanks to reducing the number of elements per equation and so the number of operations.

5.4.2 Principal Components Analysis

As shown in Section 3.1.4, this feature extraction technique is suitable for converting an original set of eventually correlated basis functions into a new uncorrelated orthogonal basis set

called principal components. For DPD purposes, PCA is typically used to reduce the number of parameters to be estimated in the feedback path and avoid ill-conditioning or over-fitting problems [Gil13]. In polynomial-based MIMO DPD architectures, a new transformed matrix for each MISO DPD block, $\hat{\boldsymbol{U}}_{\boldsymbol{i}}$ (with dimensions $L \times M_i$) is defined as,

$$\hat{\boldsymbol{U}}_{\boldsymbol{i}} = \boldsymbol{U_i V_i} \tag{5.11}$$

which corresponds to the eigenvectors of the matrix $\boldsymbol{U_i}{\boldsymbol{U_i}}^H$. The $N_i \times M_i$ transformation matrix $\boldsymbol{V_i} = (\boldsymbol{v_{i1}}, \cdots, \boldsymbol{v_{iM_i}})$ is composed of the eigenvectors of the covariance matrix ${\boldsymbol{U_i}}^H\boldsymbol{U_i}$, where N_i is the number of original basis (or DPD coefficients) of the i^{th} MISO DPD block ($i = 1, \cdots, P$), and M_i is the number of transformed basis after reduction (i.e, $M_i < N_i$). When the transformed matrix $\hat{\boldsymbol{U}}_{\boldsymbol{i}}$ is applied to (5.4), a new reduced set of transformed coefficients $\hat{\boldsymbol{w}}_{\boldsymbol{i}}$ with dimension $M_i \times 1$ is obtained. The original $N_i \times 1$ vector of coefficients is obtained through the transformation matrix as

$$\boldsymbol{w_i} = \boldsymbol{V_i}\hat{\boldsymbol{w}}_{\boldsymbol{i}}. \tag{5.12}$$

Both the original number of basis N_i and the pruning or reduction factor $RF_{di}{=}N_i/M_i$ could be different at each MISO DPD block if every PA showed different distortion characteristics. However, when the operating conditions of the PAs are similar the same parameters and reduction factor RF_d can be applied to all the MIMO branches. The transformation matrices $\boldsymbol{V_i}$ are calculated only once and will be valid irrespective of the changing transmitted waveforms when the PA operating conditions do not change very significantly over time. This process can be done *a priori* offline (i.e., factory profiling) to avoid HW resource utilization and extending the DPD adaptation time, for a given set of representative PA operating conditions. Thanks to the orthogonality of the resulting transformed matrices, the polynomial MISO DPD coefficients' extraction can be carried out with simple dot products (i.e., avoiding matrix inversions), and one-by-one increasing the number of independent components until the desired performance is reached, as shown in Chapter 3.2.

Unlike in polynomial-based approaches, where PCA is not used to directly reduce the number of the original basis functions in the DPD function, in the case of the ANN DPD, PCA does contribute to reduce the size of the ANN. When applied to the input dataset, the ANN will directly operate with the feature-reduced transformed dataset both during training and inference (or feedback and forward paths, respectively). By reducing the input dataset features and so the MIMO ANN inputs and the total amount of data to be processed by the ANN, the number of weights in the first hidden layer FN will be decreased and the overall DPD training or update time will be also reduced. Given the ANN architecture in Fig. 5.4, PCA can be applied to every MIMO baseband channel input dataset contribution $\boldsymbol{D_i}$. Therefore, considering a MIMO or MISO ANN generic input dataset $\boldsymbol{U} = (\boldsymbol{D_1}, \cdots, \boldsymbol{D_i}, \cdots, \boldsymbol{D_P})$ with $i = 1, \cdots, P$, each transformation matrix $\boldsymbol{V_i}$ can be calculated and then be applied to obtain every $\hat{\boldsymbol{D}}_{\boldsymbol{i}}$ transformed

matrix, by using the same procedure as for the polynomial MISO-DPD architecture. After applying PCA as $\boldsymbol{U^{PCA}} = \big(\text{PCA}(\boldsymbol{D_1}, RF_d), \cdots, \text{PCA}(\boldsymbol{D_P}, RF_d)\big)$, the resulting transformed input dataset will be $\hat{\boldsymbol{U}} = \boldsymbol{U^{PCA}} = \big(\hat{\boldsymbol{D}}_{\boldsymbol{1}}, \cdots, \hat{\boldsymbol{D}}_{\boldsymbol{P}}\big)$, both for the MISO (i.e., when all channels are routed into every MISO-ANN DPD block) and the MIMO ANN architectures. Applying PCA to every MIMO channel input dataset contribution works better than applying PCA to the overall multi-channel dataset since it helps to better preserve the most relevant variables of features provided by every channel. One advantage of PCA over OLS when applied to the input dataset, is that PCA is used as an unsupervised learning technique that helps extracting the hidden structure from the high dimensional dataset. PCA can produce linear combinations of the dataset features which helps the ANN reaching the desired results for a given structure.

5.5 MIMO Scenario Combined Application Procedure

This section details how the training data length reduction methods and the dimensionality reduction techniques are sequentially combined. To do that, Fig. 5.6 shows the application procedure for the MIMO DPD based on independent polynomial MISO DPD blocks (abbreviated as MISO POLY), while Fig. 5.7 and Fig. 5.8 show the procedure for the MIMO DPD with a single ANN MIMO DPD block (MIMO ANN), and the variant with independent ANN MISO DPD blocks (MISO ANN), respectively.

The flowcharts show how the PA's input-output data is employed for each architecture defined in Section 5.2, and how the polynomial behavioral modeling matrices and the ANN input datasets (also defined in Section 5.2) are processed according to the length reduction and dimensionality reduction methods deployed in Section 5.3 and Section 5.4, respectively. To reduce the complexity and processing time of the dimensionality reduction techniques the training data length reduction is first applied. In the notation being used, when a length reduction technique is applied to a variable the lr superscript is added to the variable name, while the lr_dr superscript is used when the dimensionality reduction techniques are applied to the length reduced variables.

In a first stage, if data batch selection is chosen then either SBS or CBS is applied to every $(\boldsymbol{u_i}, \boldsymbol{y_i})$ pair. The indices selection in $\boldsymbol{\beta}$ is then applied to such variables, and also to $\boldsymbol{x'_i}$ for ANN training, before building the reduced behavioral modeling and dataset matrices $\boldsymbol{U_i^{lr}}$ and $\boldsymbol{U^{lr}}$, respectively. These contain basis functions that will be a length reduced version of the original ones with L/RF_l samples each. If, otherwise, an equation selection method is chosen, the full behavioral modeling and dataset matrices, $\boldsymbol{U_i}$ and $\boldsymbol{U}$, are first created. With MeS, the method is applied P times to the $\boldsymbol{U_i}$ MISO DPD polynomial blocks or directly to the channel dataset contributions $\boldsymbol{D_i}$ in the ANNs. In either case, P equation selection index vectors $\boldsymbol{\Theta_i}$ are generated. The polynomial MIMO DPD will employ each one at its corresponding MISO-DPD block, while for the ANN dataset these index vectors will be merged into a single one by

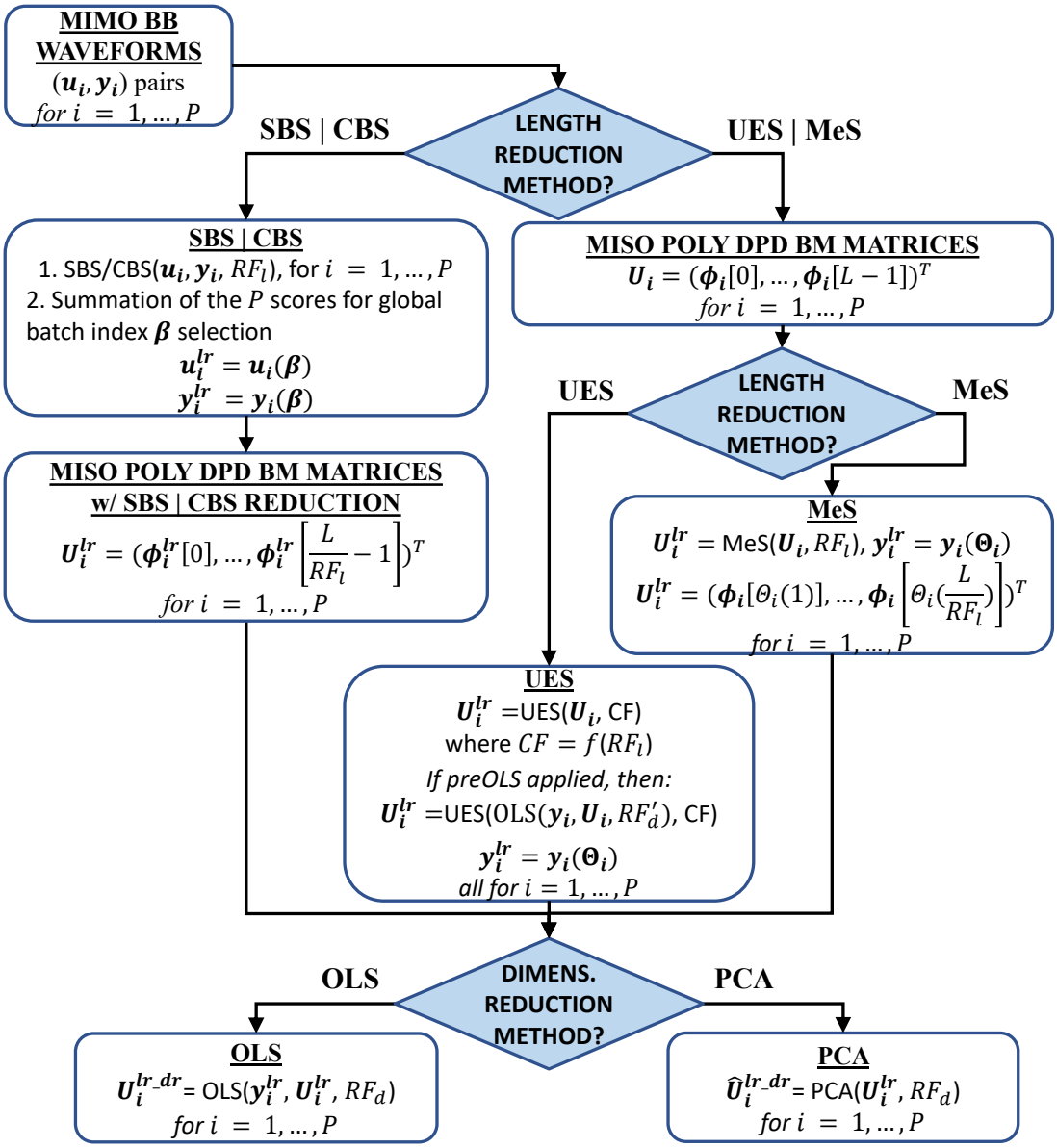


Figure 5.6: MISO POLY combined application flowchart.

following the procedure described in Section 5.3.1. As seen in Section 5.3.2, UES is applied to every MISO DPD behavioral modeling matrix $\boldsymbol{U_i}$, and just once to the $\boldsymbol{U}$ ANN dataset. An exception to the latter case is in the MISO ANN architecture when, in order to speed up UES, OLS is applied in advance to reduce the number of columns. Since OLS will be applied to $\boldsymbol{U}$ at every MISO ANN DPD block considering different target signals $\boldsymbol{x_i'}$, we can have different OLS $\boldsymbol{\Upsilon_i}$ support sets leading to different equation selections vectors $\boldsymbol{\Theta_i}$ at each of the P MISO ANN blocks. The columns selection can just be used to calculate the equation selection and not be propagated to the next stages. The reduction factor (RF_d') at the preOLS step before UES may be different to the one used for effective dimensionality reduction at the later OLS stage (RF_d).

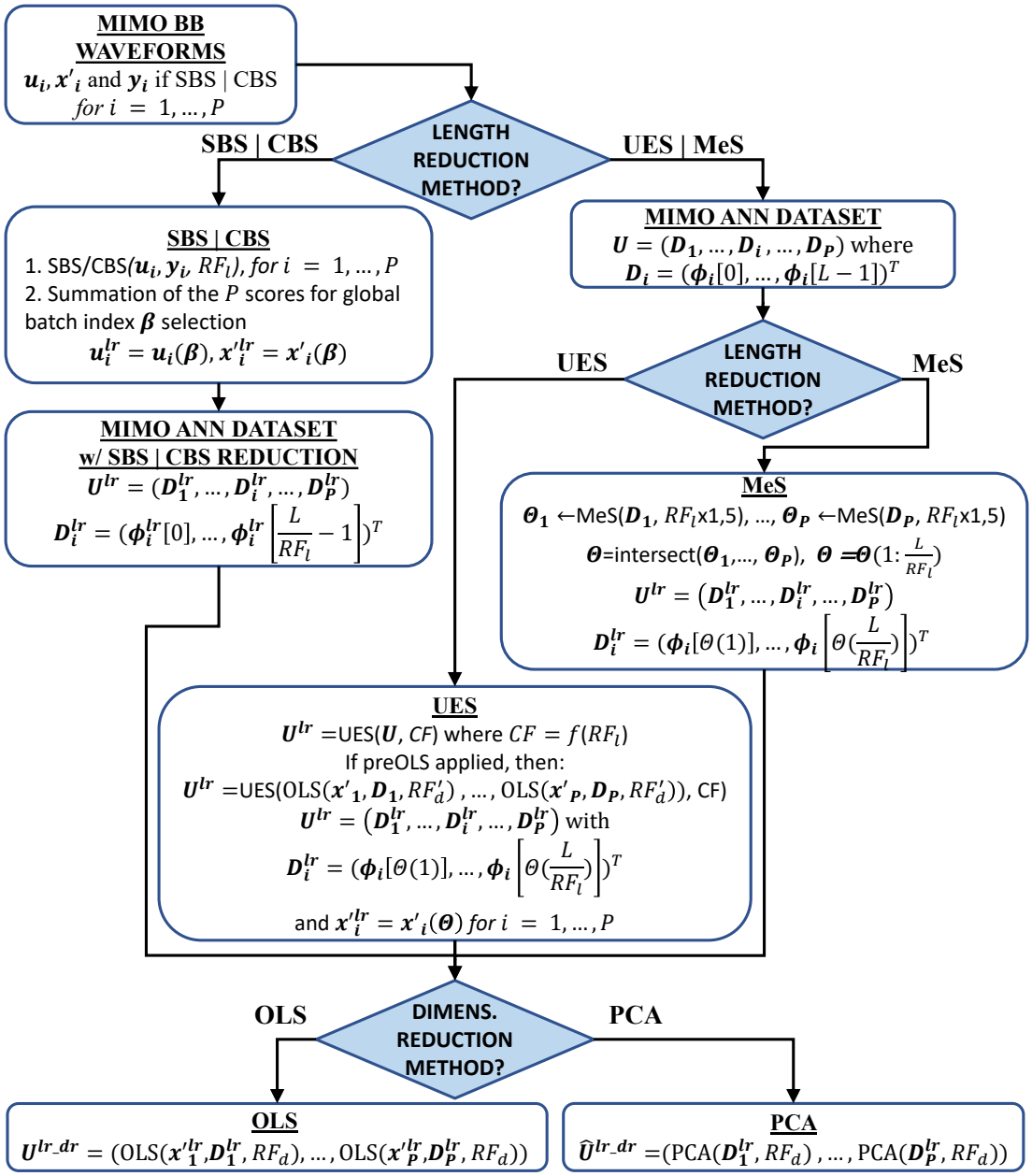


Figure 5.7: MIMO ANN combined application flowchart.

In a second stage the modeling basis or dataset variable reduction techniques OLS and PCA are applied as described in Section 5.4.1 and Section 5.4.2, respectively. The final reduced polynomial or ANN input modeling data is then obtained. As previously mentioned, both the preOLS and the OLS column indices may be calculated for every MIMO channel in advance for a given set of operating conditions to avoid increasing the MIMO DPD training/update times. Same applies to the PCA transformation matrices. Finally, it is noted that when a factor RF_d is applied in OLS or PCA to reduce all the polynomial MIMO DPD basis functions, the total number of coefficients is also reduced by this factor. In ANNs, however, this is no longer valid

since the OLS/PCA reduction is applied to the input dataset features and thus only the number of weights in the first hidden layer are reduced proportionally.

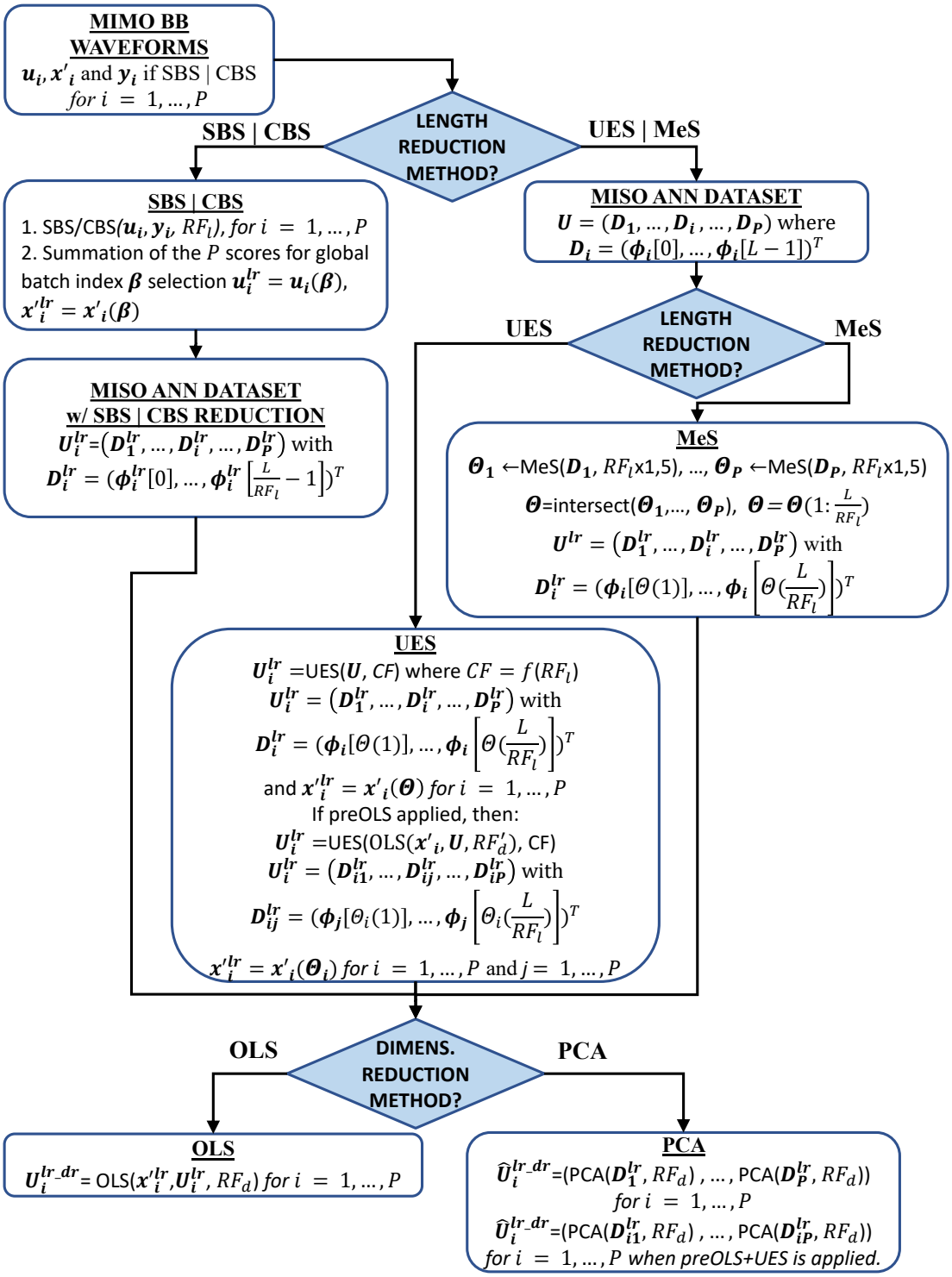


Figure 5.8: MISO ANN combined application flowchart.

5.6 Experimental Results

The proposed schemes have been benchmarked and validated with the MATLAB controlled HW test bench shown in Fig. 5.9. To launch in parallel multiple DPD experiments a MATLAB client-server waveform upload/download architecture has been enabled. The MATLAB server communicates both with the remote clients running the DPD algorithms and the laboratory HW composed of FPGA boards for waveform playback (TI TSW14J56EVM) and recording (TI TSW14J57EVM), data converters (TI DAC38RF82EVM and TI ADC12DJ5200EVM), two ADI ADL5605 PA drivers, two GaN HEMT class J PAs based on the Cree CGH35030F transistor, and a tunable bidirectional passive network to infer controlled couplings between the PA outputs and potential loading effects, (by manually setting the attenuation step of the mechanical variable attenuator shown in Fig. 5.9). The PA input cross talk effects are digitally introduced into the predistorted baseband signals, in the MATLAB Laboratory server before transmission through the test setup, according to the PA input cross talk level set by the remote user.

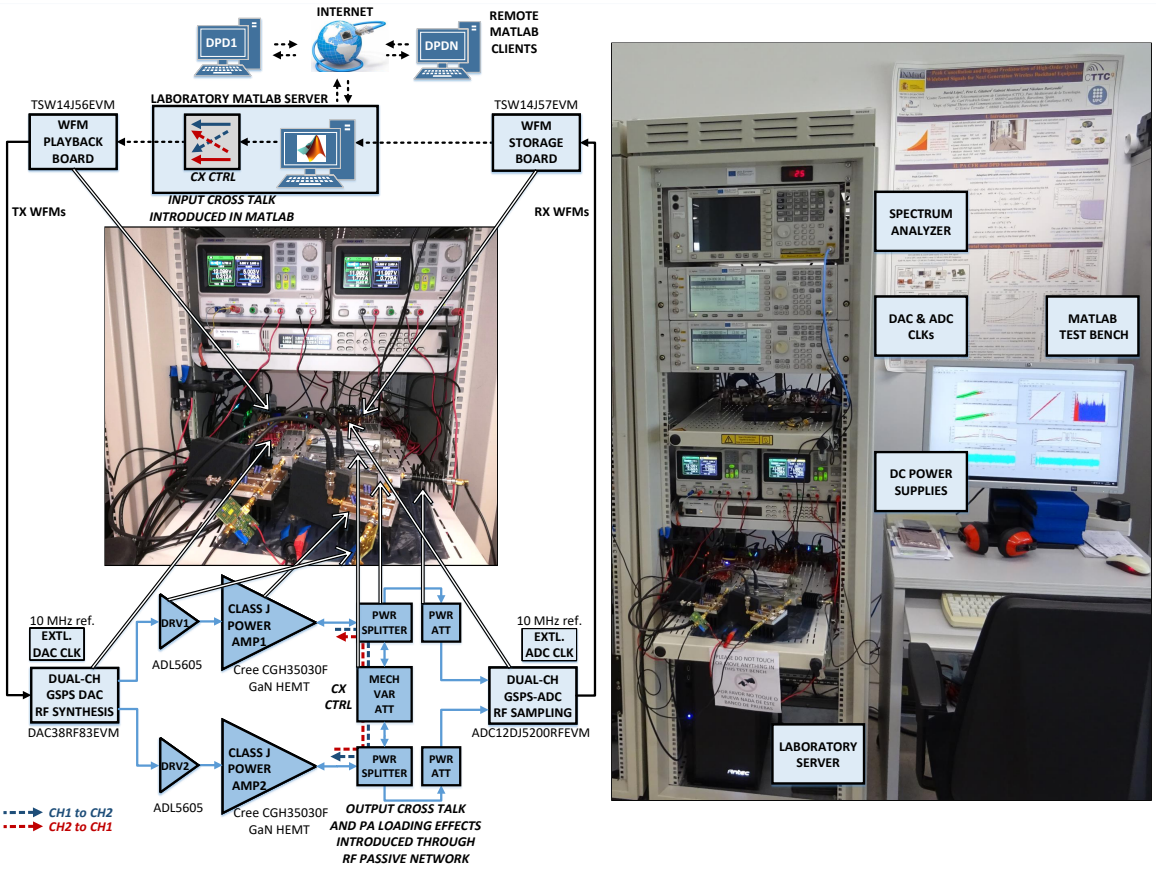


Figure 5.9: Remotely accessible multi-antenna digital linearization test setup block diagram (left) and demonstration test bench picture (right). EXTL: external; DRV: driver; CX CTRL: cross talk control; MECH VAR ATT: mechanical variable attenuator.

The PAs are operated at 875 MHz RF frequency and about 27 dBm mean output power (28V

drain-to-source voltage and around 3.1V gate-to-source voltage) with 2x2 MIMO 80 MHz bandwidth carrier-aggregated fast convolution filter bank multi-carrier (FC-FBMC) signals featuring 13-14 dB PAPR. The waveform length is 737280 samples, after oversampling 2 ms signals by a factor of 3 to accommodate for the DPD bandwidth expansion (368.64 MHz baseband sample rate). Each of the two FC-FBMC signals generated are uncorrelated and feature changing PAPR values at every training iteration and at post-training validation iterations.

The two discrete PAs need some initial adjustment to reach similar operation point due to transistor pinch-off voltage dispersion or having different parasitic effects in the two prototyped boards. The differences in linearity performance imply that there will be some common DPD modeling basis valid for the two models but to reach the maximum performance there will be basis that are needed for one of the PAs but not necessarily the other. Since the general approach is to apply the same overall DPD model to the two PAs, to reduce the DPD complexity and ensure well-conditioned identification the dimensionality reduction techniques are highly convenient.

The proposed techniques have been experimentally benchmarked and validated in Table 5.1 and Table 5.2. The parameters of the three MIMO DPD architectures evaluated are as follows:

1. **MISO POLY**. The MISO DPD block for every PA is based on the GMPNLC model with nonlinear degree 7 and up to 10-tap memory depth ($O = 4, M1 = 5, M2 = 5$). Nonlinear degree 9 and up to 12-tap memory depth was also evaluated (i.e., $O = 5, M1 = 6, M2 = 6$) but the improvement in performance was not relevant despite nearly doubling the number of coefficients.

2. **MISO ANN**. The MISO DPD blocks are based on two-hidden layer ANNs, with 20 neurons per hidden layer and hyperbolic tangent sigmoid activation functions. The following data augmentation functions are considered for every MIMO channel input dataset (I_{ini}, Q_{ini}) contribution: $|I_{ini} + jQ_{ini}|^k$ with $k = 1, \cdots, 6$, $|I_{ini}|^k$ and $|Q_{ini}|^k$ for $k = 3$, and a long-term component considering a sliding window of $K = 1000$ samples. Except for the long-term component, 12-tap consecutive memory delays are also applied to all these contributions to enable dynamic nonlinear modeling.

3. **MIMO ANN**. The single MIMO DPD block is based on a two-hidden layer ANN, with 20 neurons per hidden layer and hyperbolic tangent sigmoid activation functions. The following data augmentation functions are added to every MIMO channel input dataset (I_{ini}, Q_{ini}) contribution: $|I_{ini}+jQ_{ini}|^k$ with $k = 1, \cdots, 6$, $|I_{ini}|^k$ and $|Q_{ini}|^k$ for $k = 7, \cdots, 9$ and a long-term component considering a sliding window of $K = 1000$ samples. Again, 12-tap consecutive memory delays are applied.

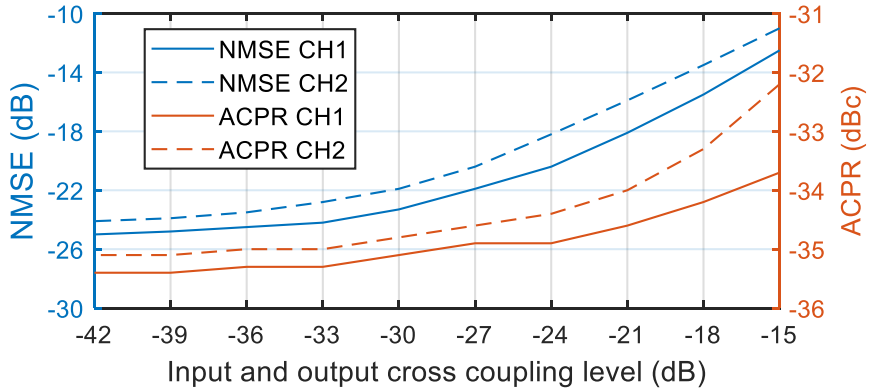


Figure 5.10: CH1 and CH2 NMSE and ACPR versus input and output cross coupling level.

These three MIMO DPD architectures and the researched techniques have been evaluated considering -15 dB PA input and -15 dB PA output cross couplings. Fig. 5.10 shows the degradation in the NMSE and ACPR figures with increasing equal input and output cross coupling values, when no MIMO DPD is applied. At -15 dB cross coupling level additional nonlinear effects appear due to changes in the input signal power statistics and the class-J PA loading conditions. At such coupling level, and considering the highest PAPR MIMO baseband signals, the PA input signal peaks are compressed about 3 dB at the PA output. No crest factor reduction techniques were applied to the MIMO waveforms. Such challenging scenario requires handling complex MIMO DPD models.

Table 5.1 and Table 5.2 show complexity, performance and timing results (i.e., overall number of MIMO DPD coefficients, NMSE and worst upper/lower ACPR, and time measured with MATLAB's tic toc functions) for a representative set of both dimensionality reduction test cases and the combination of data length and dimensionality reduction test cases, respectively [LB22b]. These techniques are evaluated with the MIMO 2x2 test setup detailed in Fig. 5.9 for the three MIMO DPD architectures. A number of test case labels are included in the tables and in the text to facilitate the tracking of the results.

Table 5.1: Dimensionality Reduction Benchmark for MIMO DPD.

	CASE. METHOD	RED. FACT.	DPD COEFF.	NMSE CH2 [dB]	NMSE CH1 [dB]	ACPR CH2 [dBc]	ACPR CH1 [dBc]	TOTAL DPD UPD. TIME [s]	1^{ST} ITER. DPD TT [s]	REMAINING ITER. TT [s]
MISO POLY GMPNLC	C1. GMP 1-D	N/A[a]	644	-11.3	-12.5	-36.7	-38.7	615		
	C2. NONE	**N/A[a]**	**1440**	**-36.3**	**-37.4**	**-46.1**	**-46.3**	**2862**		
	C3. OLS	**4**	**360**	**-36.4**	**-37.4**	**-45.4**	**-45.8**	**232**		
	C4. OLS	**8**	**180**	**-36.5**	**-37.5**	**-45**	**-45.4**	**74**		
	C5. PCA	8	180	-35.5	-36	-42.4	-43	72		
MISO ANNs	C6. OLS	6	2724	-36.3	-38.6	-46.9	-47.5	2.81E+04	1.42E+04	1.40E+04
	C7. OLS	**16**	**1604**	**-36.3**	**-36.8**	**-47.8**	**-48**	**1.75E+04**	**1.08E+04**	**6.68E+03**
	C8. PCA	**4**	**3604**	**-36.8**	**-38.8**	**-48**	**-48.7**	**3.84E+04**	**2.45E+04**	**1.38E+04**
MIMO ANN	C9. OLS	6	1764	-35.1	-37.4	-44	-45.7	4.01E+04	3.03E+04	9.82E+03
	C10. OLS	**16**	**964**	**-36.5**	**-37.6**	**-45.3**	**-46**	**1.90E+04**	**1.22E+04**	**6.83E+03**
	C11. PCA	**6**	**1764**	**-36.4**	**-38.3**	**-47.2**	**-47.7**	**5.21E+04**	**2.64E+04**	**2.58E+04**

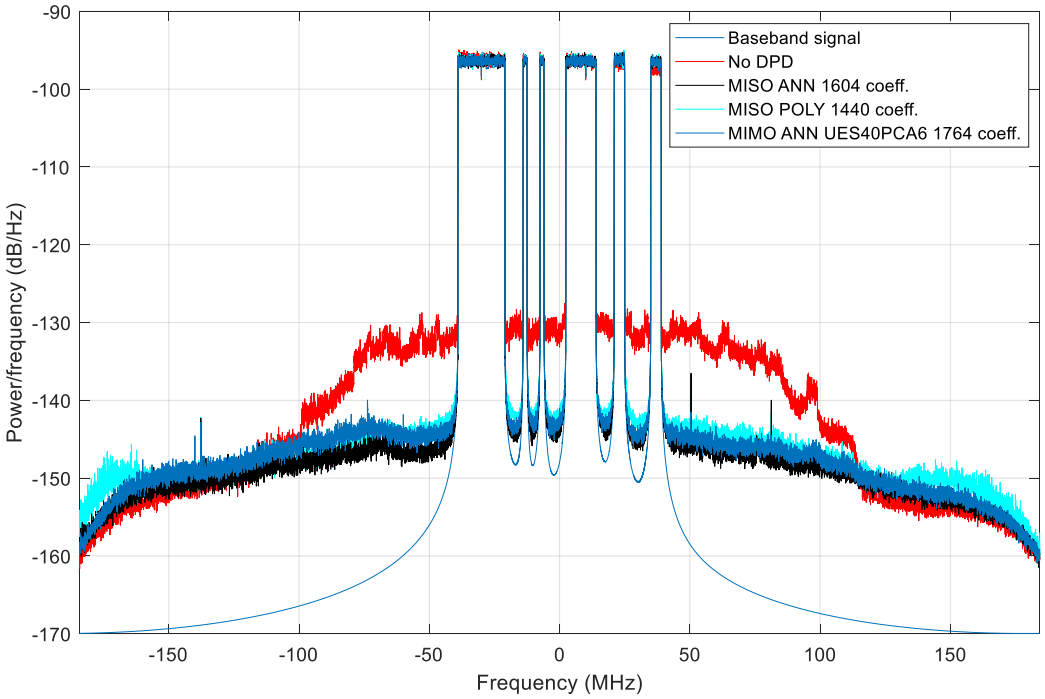
[a] Not applicable.

Regarding Table 5.1, when the best classical GMP DPD linearization configuration is applied independently to every MIMO channel (C1), since no crosstalk is accounted in the modelling, the resulting NMSE figures are equivalent to those found in Fig. 5.10 when DPD is not applied (i.e., the crosstalk is a dominant in-band effect which is not corrected, only the ACPR is enhanced by about 3-4 dB at C1). The best GMPNLC configuration (with reasonable number of coefficients) is able to overcome the previous limitations but considering the demanding test setup scenario the NMSE values reach near -37 dB NMSE and -46.2 dB ACPR on average (C2). When using the polynomial MIMO DPD scheme, OLS is the most effective technique to reduce the complexity and the total DPD update time while preserving performance. When a coefficient reduction factor of 8 is applied, OLS features about 1.5 dB and 2.5 dB better NMSE and ACPR, respectively, than PCA (C4 vs. C5). The DPD update time is reduced by a factor circa 40 when compared to the nominal case (C4 vs. C1).

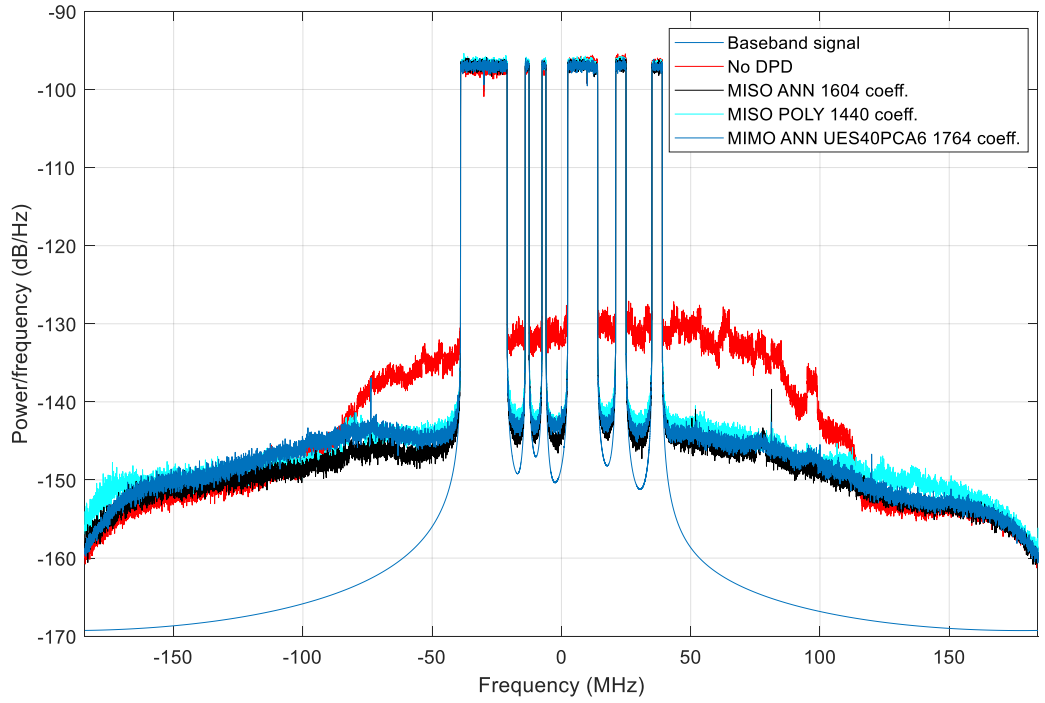
When comparing the performance of the polynomial and the ANN linearization schemes, the MISO ANN is the better performing. The results for the nominal ANN-based MIMO DPD schemes without dimensionality reduction are not provided due to the huge amount of time required to obtain them. When considering the MISO ANN scheme after applying PCA dataset reduction by 4, the NMSE is improved by 0.5-1.5 dB but, more importantly, the ACPR is also improved by 2-2.5 dB, in comparison with the GMPNLC MISO POLY scheme (C4 vs. C1). Table 5.1 also shows that both for the MISO and MIMO ANNs, for lower basis reduction factors PCA delivers the best attainable performance, but OLS is a better choice for higher reduction factors (C7 vs. C8).

When comparing the number of coefficients, for small arrays the MISO POLY scheme may feature much less coefficients than the ANN schemes (C4 vs. C10) but, however, when the best performance is pursued, they may feature similar number of coefficients. For example, the MISO ANN scheme with OLS reduction by 16 outperforms GMPNLC in terms of spectral contention as seen in Table 5.1 (C2 vs. C7) and the linearized spectra plots in Fig. 5.11. The AM-AM and the AM-PM characteristics for this case are also shown in Fig. 5.12.

The main drawback of the ANN schemes is that the DPD update time is significantly larger than the polynomial ones (i.e., 2.5 orders of magnitude higher, see C3-C4 vs C6-C11). This is clearly an obstacle for the adoption of ANN-based MIMO DPD schemes unless dimensionality reduction is combined with significant training data length reduction factors (and together with efficiently parallelized ANN-specific processing HW). The time taken to train the ANN at the first MIMO DPD iteration is in the order of magnitude (between 1 and 3 times higher) of the time needed to train the MIMO DPD in the remaining iterations with different data and PAPR statistics MIMO waveforms until the desired performance is reached (i.e., typically 7-8 iterations are needed). This is due to the fact that in the first iteration the DPD coefficients are trained from scratch, and it takes longer to find the coefficients that minimize the ANN cost function,



(a)



(b)

Figure 5.11: Power spectral density plots for (a) CH1 and (b) CH2 before and after DPD linearization, considering the GMPNLC polynomial architecture (C2), the MISO ANN architecture with OLS reduction by 16 (C7), and the MIMO ANN architecture with UES reduction by 40 and PCA reduction by 6 (C23).

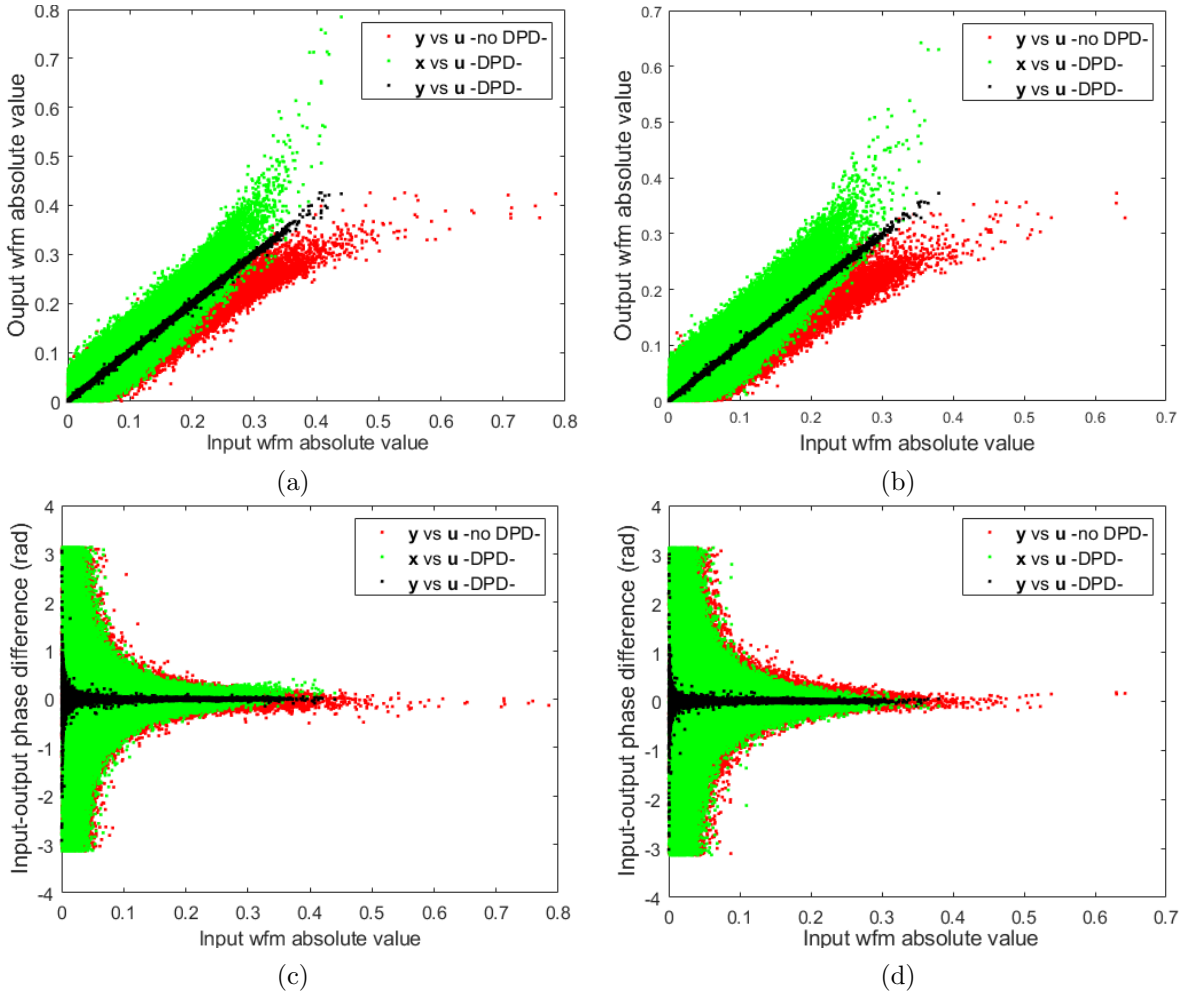


Figure 5.12: AM-AM plots for (a) CH1 and (b) CH2 and AM-PM plots for (c) CH1 and (d) CH2 before and after DPD linearization, considering the MISO ANN architecture with OLS reduction by 16 (C7).

and the following training iterations benefit from using the previously calculated and updated coefficients when training with new data. Having pre-trained ANNs can also contribute to reduce the overall MIMO DPD training time (TT in Table 5.1).

In Table 5.2, the dimensionality reduction techniques are preceded by the training data length reduction techniques as per Section 5.5, and the length reduction processing time is both identified and further added to the total DPD update time count. In the MISO POLY scheme, the best data length reduction techniques are SBS and UES, but when combined with OLS, UES performs about 1.5 dB NMSE and 1.2 dB ACPR better and is slightly faster (C14 vs. C13). MeS does not provide acceptable ACPR performance when having strong cross couplings for any of the DPD architectures (the better performing case is shown for the MISO POLY architecture in C12). In the MISO POLY case, applying OLS reduction by 8 is already a powerful tool to

Table 5.2: Training Data Length Reduction Combined with Dimensionality Reduction Benchmark for MIMO DPD.

	CASE. METHOD	RED. FACT.	DPD COEFF.	NMSE CH2 [dB]	NMSE CH1 [dB]	ACPR CH2 [dBc]	ACPR CH1 [dBc]	LENGTH REDUX TIME [s]	DPD TRAIN. TIME [s]	TOT. DPD UPD. TIME [s]
	C12. MeS-OLS	10_8	180	-32.6	-33.6	-39.7	-40.5	238.6	6.7	245.3
MISO	C13. SBS-OLS	20_8	180	-35.2	-35.7	-43.1	-43.2	50	3.3	53.3
POLY	**C14. UES-OLS**	**10_8**	**180**	**-36.8**	**-36.9**	**-44.4**	**-44.5**	**42.7**	**7.1**	**49.8**
MISO	C15. CBS-PCA	80_4	3604	-36.2	-36.8	-43.5	-43.6	68.3	146.4	214.7
ANN	**C16. UES-PCA**	**80_4**	**3604**	**-36.2**	**-37.2**	**-45**	**-44.8**	**12.9**	**92**	**104.9**
1^{st} iter	**C17. UES-PCA**	**40_4**	**3604**	**-36.8**	**-37.7**	**-45.4**	**-45.7**	**13**	**170.6**	**183.6**
pretrained	C18. UES-OLS	80_16	1604	-33.5	-35.3	-43.8	-44.5	12.9	67.5	80.4
MISO	C19. UES-PCA	80_4	3604	-35.5	-35.7	-43.1	-42.1	15.1	211.9	227
ANN	C20. UES-PCA	40_4	3604	-36.2	-37.1	-44.6	-44.6	13.3	242.3	255.6
MIMO	C21. CBS-PCA	60_6	1764	-36.4	-37.6	-45.4	-45.8	45	265	310
ANN	**C22. UES-PCA**	**60_6**	**1764**	**-36.2**	**-37.3**	**-44.9**	**-44.7**	**15.5**	**93.5**	**109**
1^{st} iter	**C23. UES-PCA**	**40_6**	**1764**	**-36.4**	**-37.6**	**-45.6**	**-45.9**	**15.2**	**200.3**	**215.5**
pretrained	C24. UES-OLS	40_16	964	-36.1	-36.6	-43.1	-43.8	15.3	87.8	103.1
MIMO	C25. UES-PCA	60_6	1764	-35.2	-36	-43	-43.1	17.7	258	275.7
ANN	C26. UES-PCA	40_6	1764	-35.9	-36.8	-45	-44.4	17.6	534.2	551.8

significantly shorten the training time. When applying UES-OLS additional 35% DPD update time reduction can be obtained paying very little performance cost with regards to applying only OLS (C4 vs. C14).

With the ANN-based schemes, when having very strong cross couplings it is hard to reach the desirable performance when applying data length reduction right at the first DPD training iteration where the ANN coefficients are calculated from scratch. To avoid NMSE and ACPR losses of about 1-1.5 dB and 2-2.5 dB (C16-C17 vs. C19-C20 and C22-23 vs. C25-C26), respectively, pretrained first iteration coefficients that have been calculated without training data length reduction can be loaded, and the length reduction be applied starting at the second iteration. This is clearly what better works for such strong crosstalk conditions but, if pretrained ANN coefficients are not available, UES can be also applied from the first DPD iteration by decreasing the UES training data length reduction factor before calculating a new set of coefficients (C16 vs. C20 and C22 vs. C26). When compared to the optimal pretrained case, the performance will be similar and the training time will be kept in the same order of magnitude (i.e., 2-4 times higher). With ANNs, and when precombined with PCA, CBS works better than SBS. While in the MISO ANN UES shows 1-1.5 dB better ACPR performance and is about 1.5 times faster than CBS (C15 vs. C16), in the MIMO ANN scheme CBS may perform similarly or even better than UES, but the DPD updating time of the former can be 2-3 times larger (C21 vs. C22). Finally, while combining UES and OLS with a high dimensionality reduction factor can be about 2 times faster than using UES-PCA with a lower dimensionality reduction factor (i.e., comparing 1^{st} iteration pretrained schemes), such advantage is not enough to compensate for the performance loss (C17 vs. C18 and C23 vs. C24).

At this point, it is important to remark that with the UES-PCA method being applied to the ANN-based MIMO DPD schemes, the polynomial MIMO DPD training times can be approached and we also have some degrees of freedom to overcome the linearization performance of the polynomial. One example may be found in Table 5.2 by comparing the MISO POLY case when

applying OLS reduction by 4 with either the MISO ANN or the MIMO ANN cases applying UES-PCA (C4 vs. C17 and C23). The linearized spectra plots obtained when using the MIMO ANN scheme after applying UES-PCA in C23 are also added to Fig. 5.11 to show similar spectral contention when compared to the MISO POLY case without basis reduction in C2 (i.e., the comparison is done here for a similar number of coefficients, but featuring the MIMO-ANN scheme lower training time). The total DPD update time of the MISO/MIMO ANN schemes is reduced about 2 orders of magnitude when comparing the use of PCA or OLS with either UES-PCA or UES-OLS (C6-C11 vs. C15-C26). The gain with respect to the nominal ANNs without dimensionality reduction would be above three orders of magnitude. Applying dimensionality reduction to the herein ANN topologies (for 2x2 MIMO DPD) with factors between 6 and 16 deliver reductions in the number of ANN coefficients between the 55% and the 75%, respectively. Note that the higher the number of antennas, the more impact that the dimensionality reduction will have over the total number of ANN coefficients (the weights at the first hidden layer will increase far more significantly than the weights and biases at others).

The most representative results combining different MIMO DPD architectures with the proposed techniques are also summarized in Fig. 5.13. In general terms, it is evident that according to our 2x2 test scenario the ANN-based DPD schemes may feature higher complexity than the polynomial ones, but also similar adaptivity speed and better performance. Such benefits could even be more evident when considering larger arrays where the number of MIMO DPD coefficients may increase with the number of antennas much faster in the polynomial approach than in ANNs (i.e., MIMO ANN scheme).

The experimental benchmarking also highlights the advantages and disadvantages of applying each of the three MIMO DPD architectures. Despite the reach of the measurements is limited in number of antennas, several projections regarding the application of the MIMO DPD

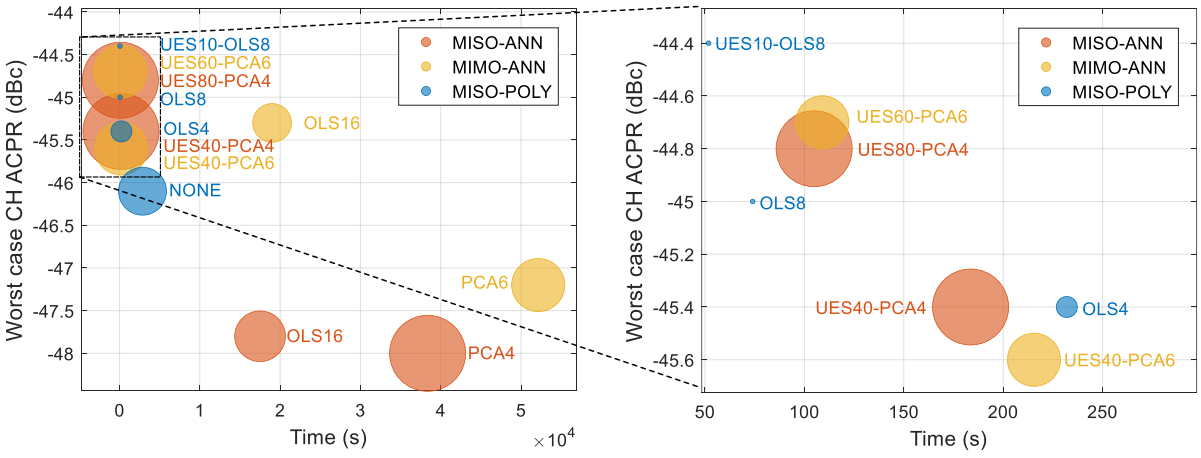


Figure 5.13: Bubble plot with time versus worst case channel ACPR and number of coefficients (i.e., proportional to bubble size) considering the three presented MIMO DPD architectures.

architectures to large arrays can be envisaged from the results obtained. For a small number of antennas, the polynomial architecture is valid if the nonlinear scenario is not very complex. The implementation is the most affordable with dimensionality reduction and the adaptation will be the fastest. For large arrays the polynomial MISO DPD approach may be hard to implement unless other schemes considering dual-input DPD architectures [Hau18] are exploited. When either the performance or the DPD updating time are prominent requirements, using multiple MISO DPD ANNs is the best choice. This scheme will have more coefficients than the fully MIMO neural network but dividing the problem into smaller ones contributes to both featuring better performance and improve the adaptation time of the latter. Replicating such networks for large arrays, however, could be resource consuming. Thus, a fully MIMO ANN architecture may be a more affordable choice for scenarios with PAs offering similar distortion and not having the worst nonlinear effects since in this scheme the cost function being minimized accounts for the sum of the errors at each transmitter channel. Finally, another option that must be considered, on top of the ones evaluated in this section, is the use of independent MIMO DPD ANNs having more inputs than outputs and being aware of the radio HW and antennas spatial arrangement to get optimal MIMO DPD block input baseband channel routings (i.e., it is an hybrid approach between the single-ANN MIMO DPD scheme and the one built with independent MISO DPD ANNs).

Chapter 6

Conclusion

6.1 Main Findings

The main research work conducted in the thesis can be seen as a ML toolkit of techniques able to reduce complexity and speed up DPD training, while guaranteeing robustness and performance requirements like the NMSE, EVM or ACPR. These techniques are versatile enough for use in SISO and MIMO polynomial and ANN DPD systems, and can be applied to sub-7 GHz systems and, potentially, to mm-wave systems given the baseband operation of the adaptive DPD system. Every contribution has been experimentally validated and the benefits are summarized in Figure 6.1, according to the application scenarios. The findings linked to the main research contributions are detailed as follows:

In **polynomial SISO-DPD**, **the independent PCA-DPD identification method** leverages on using the PCA theory to find a transformed matrix with new orthogonal components. With the independent DPD identification it is possible to trade-off updating or identification convergence time versus computational complexity. That is, updating in parallel at every iteration (in a more relaxed timescale than the real-time DPD in the forward path), following a LMS gradient technique, as many components as desired (trade-off convergence time vs. FPGA resources). The independent DPD relies on the need to find a transformation matrix, which thanks to the APCA algorithm can be iteratively found and, moreover, it is suitable to be implemented online in an FPGA. Therefore, it has been proved a new identification algorithm by merging the APCA technique with the closed-loop DPD adaptive estimation. This way, a robust independent DPD identification is obtained by using the minimum necessary coefficients in the adaptation subsystem to meet specific linearity levels.

As an alternative to polynomial SISO-DPD based models, some design guidelines for **ANN SISO-DPD** approaches to model and compensate for the nonlinear distortion introduced by the PA or other unwanted distortion effects such as IQ imbalances and dc offset, have been

ML TOOLKIT FOR ADAPTIVE DIGITAL PREDISTORTERS	- OBJECTIVE 1 - PERFORMANCE & ROBUSTNESS	- OBJECTIVE 2 - DPD COMPLEXITY REDUCTION	- OBJECTIVE 3 - SPEED UP DPD UPDATING
SISO POLYNOMIAL DPD			
• CFR + PIECEWISE & LT DPD & MeS FOR 5G TDD	• Supports burst-like wfm, wrt GMP-DPD, up to -7 dB NMSE and -9 dB in ACPR in worst PAPR waveforms, max. var. in dB for NMSE and ACPR (same PAPR) cut to half.	• N/A but potential ADC sample rate reduction	• N/A
• CFR + GMP-DPD FOR BB-PPDR SHARED SPECTR.	• Wrt NO DPD, 2 dB higher Pout, +12 drain eff. incr. (%), -17 dBc in ACPR, -12 dB in EVM, up to 18 dB in SINR	• N/A	• N/A
• GMP-DPD SW-OMP FOR 4G CA	• Wrt OMP, - 2 dB in ACPR	• x3 reduction in coeffs. for DPD estimation/adaptation	• Not evaluated
• CFR + GMP-DPD & PCA FOR BACKHAUL RADIOS	• Wrt NO DPD, 2 dB higher Pout, - 5 dB in ACPR, -9.5 dB in EVM	• x3 reduction in coeffs. for DPD estimation/adaptation	• Not evaluated
• GMP-DPD INDEPENDENT PCA ESTIMATION	• Supports independent incremental DPD coeff. estimation until linearity is met, wrt NO DPD, -20 dB in NMSE, -9 dB in ACPR, -12.5 dB in EVM	• x3.5 reduction in coeffs. for DPD estimation/adaptation	• Not evaluated
SISO ANN DPD			
• CBS/SBS-DOMP ANN DPD	• Similar performance as GMP DPD when update time close	• x2.5 reduction in ANN DPD coeffs.	• 8-iteration ANN DPD update time reduced by about 75
MIMO POLYNOMIAL DPD			
• UES-OLS FOR GMPNLC MIMO DPD	• Performance requirements fulfilled	• Up to x10 reduction in GMPNLC DPD coeffs.	• 8-iteration GMPNLC DPD update time reduced by about 60
MIMO ANN DPD			
• UES-PCA FOR ANN MIMO DPD	• Performance requirements fulfilled	• Up to x2 reduction in ANN DPD coeffs. Potential reduction scalability	• 8-iteration ANN DPD update time reduced by about 230

Figure 6.1: Summary of the benefits brought by the techniques researched in the thesis (light and dark grey colors for early works and main contributions, respectively).

presented. Some pros and cons that need to be taken into account in order to decide whether to use the solutions based on ANNs are discussed in the following. While ANN-based models permit a compact representation of a multidimensional function, they suffer from two main weaknesses: a) ANNs often converge on local minima rather than global minima and b) can suffer from the over-fitting problem due to the traditional empirical risk minimization principle employed by ANNs. Regarding the dimensions, ANNs are parametric models whose size is fixed and depends on their specific architecture. Under the current practical circumstance, ANNs have been successfully employed in several applications in the DFE for their capability to approximate any continuous function. It is strongly believed that ANN based solutions will play a very important role in future DFE implementation and deployment including intelligent vector processors and

functional intellectual property (IP) blocks. With respect to SISO ANN DPD, it has been shown that the direct learning ANN DPD can outperform the classical DPD schemes at the cost of a potentially unaffordable training time for adaptive scenarios. The DOMP algorithm can be combined with batch selection input dataset length reduction strategies. The **CBS/SBS-DOMP** scheme is effective for reducing the number of ANN coefficients and to greatly cut the ANN training time for SISO DPD adaptation while fulfilling signal quality requirements.

The **polynomial and ANN MIMO-DPD** work has provided an insight on training data selection and dimensionality reduction techniques for faster adaptivity and more efficient MIMO predistorters. The MIMO ANN DPD schemes can outperform the polynomial ones in complex scenarios but show unaffordable training times. The number of parameters required for DPD or PA behavioral modeling in MIMO systems based on parametric approaches can easily grow exponentially, which increases the computational complexity and may provoke over-fitting and uncertainty in the estimation. Feature selection and feature extraction dimensionality reduction techniques oriented at reducing the number of required parameters and guaranteeing a well-conditioned extraction have been adopted into such scenarios. OLS and PCA allow reducing the number of DPD coefficients and training time but need to be combined with data length reduction techniques for fast adaptivity. The batch selection techniques and the UES method contribute to significantly reduce the ANN DPD training time and overcome the limitations of histogram-based data selection mechanisms under strong cross couplings. This work contributes with two novel MIMO-DPD strategies: the first applies **UES-OLS** for polynomial predistorters, and the second **UES-PCA** for ANN predistorters.

6.2 Future Work

There are several considerations to keep in mind as a continuation of the thesis work and for implementation of the DPD techniques researched in either forthcoming 5G radio equipment or its evolution towards 6G. These are listed as follows:

- **MIMO DPD**: The UES-OLS and UES-PCA techniques need to be experimentally evaluated for sub-7 GHz massive MIMO and mm-wave hybrid beamforming transceivers with larger number of antennas and be properly quantified in terms of digital implementation resource occupation both for polynomial and ANN predistorters. When thinking in implementation, more effective algorithmic versions of OLS and PCA need to be considered to support the plethora of different fast-changing scenarios that may occur in 5G and beyond systems. Also, either piece-wise or decision-tree based model switching could be integrated into the proposed MIMO-DPD schemes to fasten DPD adaptation. Finally, when targeting mm-wave hybrid beamformers with higher signal bandwidth than sub-7 GHz systems, and where having a dedicated DPD observation receiver is a harder design decision, the thesis

MIMO-DPD work needs also to be applied to systems with either a switched single conducted DPD receiver, a local over-the-air (OTA) DPD receiver o a remote DPD receiver at user equipment, where also the radio channel needs to be accounted and feedback between the UE and the BS is required.

- **ANN DPD**: Nowadays, the ANN are seen as a potential enhancer for polynomial DPDs but hardly seen as a replacement due to their complexity and training times. There are still many unknowns with regard to the replacement of classical PHY functions by AI blocks or even new end-to-end Tx-Rx AI-native systems, but very likely the ANNs will be gaining momentum over the following years. So, depending on the application scenario, other ANN architectures like convolutional neural networks (CNN) or long short-term memory (LSTM) architectures (i.e., for long term memory modeling) may be required. In both cases, the outcomes of this work are fully applicable to shorten the training time and complexity of the current and future MIMO DPD systems. But additionally, it will be convenient to continue exploring techniques to adaptively prune not only the dataset but the inner ANN architecture depending on the current radio scenario. Finally, continual learning ANN models accounting for the waveform or PA operation conditions as ANN inputs, or federated learning models to leverage on the learning of physically distributed radios and cover a higher number of different waveform and PA operation cases in shorter time by utilizing a centralized model. The latter can be an interesting approach when a vast deployment of multi-antenna small cells radios is done with similar radios (i.e., same radio equipment manufacturer).

- **HYBRID DPD**: The incessant user capacity demands can only be met in next generation communication equipment by shifting its operation to mm-wave and beyond mm-wave frequencies with higher bandwidth availability. When thinking of transmitter arrays with massive bandwidths and moderate or low PA power levels, and relaxed ACPR requirements (as in 5G mm-wave operation) the DPD approach can hardly fit into such scenario. Very likely, hybrid DPD approaches (HPD) will be needed: The idea is to linearize the PA's static nonlinear behavior using a digitally-controlled analog predistorter (APD) before the PA together with an adaptive DPD operating cooperatively in a concurrent way. The APD will be responsible for cancelling the static in-band and out-of-band nonlinearities. Furthermore, the DPD will cope with the cancellation of the in-band dynamic nonlinearities due to the PA memory effects, thus avoiding the need for a higher sampling rate at baseband.

- **C-RAN MEC DPD**: The beyond 5G networks will combine high capacity and flexible cloud radio access network architectures (C-RAN) with mobile edge computing (MEC) to enable lower end-to-end latencies. Such approach potentially allows for new intelligent DPD processing paradigms distributing the processing load between the radio unit (RU)

and the MEC platform that equips higher-end dedicated hardware and more energy sustainable power supply. This may contribute to i) reducing the DPD adaptation time by one additional order of magnitude and thus shorten the gap to reach millisecond-level DPD update in highly demanding systems, ii) to significantly reduce the RU DPD processing resources in the digital front-end, iii) enhance the overall radio sustainability. Another option to be considered is the use of radio-as-a-service function deployment splitting the DPD computation among heterogeneous edge-to-radio HW, or replacing the whole DPD system in the FPGA through partial reconfiguration techniques, considering higher or lower complexity models according to the KPIs that need to be met at every time. Such reconfiguration can be based on the traffic information, number of users, etc., but also based on different sensors accounting for environmental factors or visualizing in more detail which is the radio communication scenario according to the users' physical distribution. Finally, the use of HPD systems could be seen as one of the enablers of the C-RAN MEC MIMO-DPD approach by combining the in-band distributed DPD processing (with significantly less information to be shared between RAN entities) and AI techniques able to set the optimal digitally tuned parameters of the analog predistorter focusing on the out-of-band distortion.

The years to come will be more challenging than ever for the DPD system but, at the same time, the multidisciplinary approach that has been manifested in this partial future work vision makes also the challenge more exciting than ever.

Bibliography

[Abd16] A. Abdelhafiz, L. Behjat, F. M. Ghannouchi, M. Helaoui, and O. Hammi, "A high-performance complexity reduced behavioral model and digital predistorter for MIMO systems with crosstalk", *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1996–2004, May 2016.

[Alv20] L. Alvarez-López, J. A. Becerra, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Determining a digital predistorter model structure for wideband power amplifiers through random forest", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, pp. 50–52, 2020.

[Ami14] S. Amin, P. N. Landin, P. Händel, and D. Rönnow, "Behavioral modeling and linearization of crosstalk and memory effects in RF MIMO transmitters", *IEEE Trans. Microw. Theory Techn.*, vol. 62, no. 4, pp. 810–823, Apr. 2014.

[Ant10] L. Anttila, P. Händel, and M. Valkama, "Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters", *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 730–739, Apr. 2010.

[Aue10] G. Auer, O. Blume, V. Giannini, I. Godor, M. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, *et al.*, "D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown", *Earth*, vol. 20, no. 10, Dec. 2010.

[Bar02] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face recognition by independent component analysis", *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, Nov. 2002.

[Bar16] T. Barton, "Not just a phase: Outphasing power amplifiers", *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 18–31, Feb. 2016.

[Bar17a] R. M. Barnett, "Principal component analysis", GeostatisticsLessons.com. [Online]. Available: https://geostatisticslessons.com/pdfs/principalcomponentanalysis.pdf (accessed May, 2023), Dec. 2017.

[Bar17b] N. Bartzoudis, O. Font-Bach, M. Miozzo, C. Donato, P. Harbanau, M. Requena, D. López, I. Ucar, A. Azcorra Saloña, P. Serrano, J. Mangues, and M. Payaró, "Energy footprint reduction in 5g reconfigurable hotspots via function partitioning and bandwidth adaptation", *5th Int. Workshop Cloud Technol. Energy Efficiency in Mobile Commun. Netw.*, pp. 1–6, 2017.

[Bar21] 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, pp. 5772, Aug. 2021.

[Bas09] S.A. Bassam, M. Helaoui, and F. M. Ghannouchi, "Crossover digital predistorter for the compensation of crosstalk and nonlinearity in MIMO transmitters", *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1119–1128, May 2009.

[Bäu96] R. W. Bäuml, R. F. H. Fischer, and J. B. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping", *Electron. Lett.*, vol. 32, no. 22, pp. 2056, Oct. 1996.

[Bec18a] J. A. Becerra, X. Herrera, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Sparse model selection of digital predistorters using subspace pursuit", *in Proc. Eur. Microw. Integr. Circ. Conf.*, pp. 190–193, 2018.

[Bec18b] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. García-Frías, and G. Arce, "A doubly orthogonal matching pursuit algorithm for sparse predistortion of power amplifiers", *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 8, pp. 726–728, Aug. 2018.

[Ber96] D. P. Bertsekas, and J. N. Tsitsiklis, "Neuro-dynamic programming", *Encyclopedia of Optimization*, pp. 2555–2560, Springer US, 1996.

[Bla34] H. S. Black, "Stabilized feedback amplifiers", *Bell Syst. Techn. J.*, 1934.

[Blu07] T. Blumensath, and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares", Project report, Mar. 2007.

[Boo11] H. H. Boo, S. Chung, and J. L. Dawson, "Digitally assisted feedforward compensation of cartesian-feedback power-amplifier systems", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 58, no. 8, pp. 457–461, Aug. 2011.

[Bra08] R.N. Braithwaite, "Wide bandwidth adaptive digital predistortion of power amplifiers using reduced order memory correction", *in IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 1517–1520, 2008.

[Bra11] R. N. Braithwaite, "General principles and design overview of digital pre-distortion", *Digital Front-End in Wireless Communications and Broadcasting*, F. Luo, Ed. Cambridge, U.K.: Cambridge Univ. Press, pp. 143–191, 2011.

[Bra12] R. N. Braithwaite, "Reducing estimator biases due to equalization errors in adaptive digital predistortion systems for RF power amplifiers", *in IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 1–3, June 2012.

[Bra15] R. Neil Braithwaite, "A comparison of indirect learning and closed loop estimators used in digital predistortion of power amplifiers", *2015 IEEE MTT-S Int. Microw. Symp.*, IEEE, May 2015.

[Bra17] R. N. Braithwaite, "Digital predistortion of an RF power amplifier using a reduced Volterra series model with a memory polynomial estimator", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3613–3623, Oct. 2017.

[Bri20a] A. Brihuega, M. Abdelaziz, L. Anttila, M. Turunen, M. Allen, T. Eriksson, and M. Valkama, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements", *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, Sept. 2020.

[Bri20b] A. Brihuega, L. Anttila, M. Abdelaziz, T. Eriksson, F. Tufvesson, and M. Valkama, "Digital predistortion for multiuser hybrid MIMO at mmWaves", *IEEE Trans. Signal Process.*, vol. 68, pp. 3603–3618, 2020.

[Bri20c] A. Brihuega, L. Anttila, and M. Valkama, "Neural-network-based digital predistortion for active antenna arrays under load modulation", *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 8, pp. 843–846, Aug. 2020.

[CC17] J. Chani-Cahuana, M. Ozen, C. Fager, and T. Eriksson, "Digital predistortion parameter identification for RF power amplifiers using real-valued output data", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 10, pp. 1227–1231, Oct. 2017.

[Ces07] A. Cesari, J. M. Dilhac, P. L. Gilabert, G. Montoro, and E. Bertran, "A FPGA-based platform for fast prototyping of RF PA predistortion linearizers", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, 2007.

[Che89] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification", *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[Che90] D. L. Chester, "Why two hidden layers are better than one", *in Proc. International Joint Conference on Neural Networks, Washington, D. C.*, vol. 1, pp. 265–268, 1990.

[Che91] S. Chen, C.F.N. Cowan, and P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks", *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 302–309, Mar. 1991.

[Che02] T. Chen, Y. Goren, C. Jensen, P. Lally, and D. Gagne, "A novel technology for linearizing traveling wave tube amplifiers", *in IEEE MTT-S Int. Microw. Symp. Dig. (Cat. No.02CH37278)*, vol. 2, pp. 773–776 vol.2, 2002.

[Chi35] H. Chireix, "High power outphasing modulation", *in Proc. Inst. Radio Eng.*, vol. 23, no. 11, pp. 1370–1392, Nov. 1935.

[Cox74] D. Cox, "Linear amplification with nonlinear components", *IEEE Trans. on Commun.*, vol. 22, no. 12, pp. 1942–1945, Dec. 1974.

[Cri06] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, Artech House, Inc., Norwood, MA, USA, 2006.

[Cro23] L. H. Crockett, D. Northcote, and R. W. Stewart (eds.), *Software Defined Radio with Zynq Ultrascale+ RFSoC*, Strathclyde Academic Media, Jan. 2023.

[Cyb89] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Math. Control Signals Syst.*, vol. 2, no. 4, pp. 303–314, Dec. 1989.

[Deb11] B. Debaillie, A. Giry, J. M. Gonzalez, L. Dussopt, M. Li, D. Ferling, and V. Giannini, "Opportunities for energy savings in pico/femto-cell base-stations", *Future Netw. Mobile Summit*, pp. 1–8, 2011.

[Doh36] W. H. Doherty, "A new high efficiency power amplifier for modulated waves", *in Proc. Inst. Radio Eng.*, vol. 24, no. 9, pp. 1163–1182, 1936.

[Due17] Jed A. Duersch, and Ming Gu, "Randomized qr with column pivoting", *SIAM J. on Scient. Comput.*, vol. 39, no. 4, pp. C263–C291, 2017.

[Eri22] Telefonaktiebolaget LM Ericsson, "Ericsson Mobility Report November 2022", Ericsson Reports and Papers. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/november-2022 (accessed May, 2023), Nov. 2022.

[Fag19] C. Fager, T. Eriksson, F. Barradas, K. Hausmair, T. Cunha, and J. C. Pedro, "Linearity and efficiency in 5G transmitters: New techniques for analyzing efficiency, linearity, and linearization in a 5G active antenna transmitter context", *IEEE Microw. Mag.*, vol. 20, no. 5, pp. 35–49, May 2019.

[FB15] O. Font-Bach, N. Bartzoudis, X. Mestre, D. López-Bueno, P. Mege, L. Martinod, V. Ringset, and T. A. Myrvoll, "When SDR meets a 5G candidate waveform : Agile

use of fragmented spectrum and interference protection in PMR networks", *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 56–66, Dec. 2015.

[FB17] O. Font-Bach, N. Bartzoudis, D. López-Bueno, X. Mestre, P. Mège, L. Martinod, T. A. Myrvoll, and V. Ringset, "Chapter 20 - real-time implementation and experimental validation of a DL FBMC system", Markku Renfors, Xavier Mestre, Eleftherios Kofidis, Faouzi Bader (eds.), *Orthogonal Waveforms and Filter Banks for Future Communication Systems*, pp. 513–546, Academic Press, 2017.

[For15] Small Cell Forum, "Backhaul for Urban Small Cells. A Topic Brief.", SCF Publications. [Online]. Available: https://scf.io/en/documents/095_-_Backhaul_for_urban_small_cells_a_topic_brief.php (accessed May, 2023), June 2015.

[Fra98] S. Fragiacomo, C. Matrakidis, and J.J. O'Reilly, "Multicarrier transmission peak-to-average power reduction using simple block code", *Electron. Lett.*, vol. 34, no. 10, pp. 953, May 1998.

[Fre21] C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. S. Blair, and A. Friday, "The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations", *Patterns*, vol. 2, no. 9, pp. 100340, 2021.

[Gao18] W. Gao, "Linearization of wideband Wi-Fi power amplifiers using RF analog memory predistortion", *IEEE Int. Conf. Commun.*, pp. 1-6, 2018.

[Gem92] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma", *Neural Comput.*, vol. 4, no. 1, pp. 1–58, Jan. 1992.

[Gha13] H. Ghandi, D. Greenstreet, and J. Quintal, "Digital radio front-end strategies provide game changing benefits for small cell base station", Texas Instruments White Paper. [Online]. Available: http://www.ti.com/lit/wp/spry236/spry236.pdf [Accessed July, 2016], May 2013.

[Gil04] P. L. Gilabert, E. Bertran, G. Montoro, and J. Berenguer, "Study on the robustness of a 22 MHz bandwidth feedforward amplifier at the 2.4 GHz ISM-band", *in Proc. IEEE Pers., Indoor Mobile Radio Commun.*, vol. 1, pp. 186–190, Sept. 2004.

[Gil07a] P. L. Gilabert, *Multi lookup table digital predistortion for RF power amplifier linearization*, PhD Thesis, Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 2007.

[Gil07b] P.L. Gilabert, G. Montoro, and E. Bertran, "An Overview on Digital Baseband Predistortion for Linearising RF Power Amplifiers", *European Microw. Week Workshop on Wireless Transmitter Design for Future Wireless Syst.*, 2007.

[Gil12] P.L. Gilabert, and G. Montoro, "Look-up table implementation of a slow envelope dependent digital predistorter for envelope tracking power amplifiers", *IEEE Microw. Wireless Compon. Lett.*, vol. 22, no. 2, pp. 97–99, Feb. 2012.

[Gil13] P. L. Gilabert, G. Montoro, D. López, Ni. Bartzoudis, E. Bertran, M. Payaró, and A. Hourtane, "Order reduction of wideband digital predistorters using principal component analysis", *in IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 1–4, 2013.

[Gil15] P. L. Gilabert, and G. Montoro, "3-D Distributed memory polynomial behavioral model for concurrent dual-band envelope tracking power amplifier linearization", *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 638–648, Feb. 2015.

[Gil16] P. L. Gilabert, G. Montoro, *et al.*, "Comparison of model order reduction techniques for digital predistortion of power amplifiers", *in Proc. Eur. Microw. Conf.*, pp. 182–185, Oct. 2016.

[Gil18] P. L. Gilabert, D. López-Bueno, and G. Montoro, "Spectral weighting orthogonal matching pursuit algorithm for enhanced out-of-band digital predistortion linearization", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, pp. 1–1, 2018.

[Gil19] P. L. Gilabert, G. Montoro, D. Vegas, N. Ruiz, and J. A. García, "Digital predistorters go multidimensional: DPD for concurrent multiband envelope tracking and outphasing power amplifiers", *IEEE Microw. Mag.*, vol. 20, no. 5, pp. 50–61, May 2019.

[Gil20a] P. L. Gilabert, R. N. Braithwaite, and G. Montoro, "Beyond the moore-penrose inverse: Strategies for the estimation of digital predistortion linearization parameters", *IEEE Microw. Mag.*, vol. 21, no. 12, pp. 34–46, Dec. 2020.

[Gil20b] P. L. Gilabert, D. López-Bueno, Thi Q. A. Pham, and G. Montoro, "Machine Learning for Digital Front-End", *Machine Learning for Future Wireless Communications*, Chap. 17, pp. 327–381, John Wiley & Sons, Ltd, 2020.

[Gil20c] P. L. Gilabert, D. Vegas, Z. Ren, G. Montoro, J. R. Pérez-Cisneros, M. N. Ruiz, X. Si, and J. A. García, "Design and digital predistortion linearization of a wideband outphasing amplifier supporting 200 MHz bandwidth", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, pp. 46–49, 2020.

[Gua12] L. Guan, and A. Zhu, "Optimized low-complexity implementation of least squares based model extraction for digital predistortion of RF power amplifiers", *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 594–603, Mar. 2012.

[Gue17] R. Guerzaguet, N. Bartzoudis, L. Gomes Baltar, V. Berg, J.-B. Doré, D. Kténas, O. Font-Bach, X. Mestre, M. Payaró, M. Färber, and K. Roth, "The 5G candidate

waveform race: a comparison of complexity and performance", *EURASIP J. Wireless Commun. and Netw.*, vol. 2017, no. 13, Jan. 2017.

[Gui22] E. Guillena, W. Li, G. Montoro, R. Quaglia, and P. L. Gilabert, "Reconfigurable DPD based on ANNs for wideband load modulated balanced amplifiers under dynamic operation from 1.8 to 2.4 GHz", *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 1, pp. 453–465, Jan. 2022.

[Gum18] K. Gumber, and M. Rawat, "Low-cost RFin-RFout predistorter linearizer for high-power amplifiers and ultra-wideband signals", *IEEE Trans. Instrum. Meas*, vol. 67, no. 9, pp. 2069–2081, Sept. 2018.

[Hag94] M.T. Hagan, and M.B. Menhaj, "Training feedforward networks with the Marquardt algorithm", *IEEE Trans. Neural Netw.*, vol. 5, no. 6, pp. 989–993, Nov. 1994.

[Hag96] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*, PWS Publishing, Boston, 1996.

[Han05] S. H. Han, and J. H. Lee, "Modulation, coding and signal processing for wireless communications - an overview of peak-to-average power ratio reduction techniques for multicarrier transmission", *IEEE Wireless Commun.*, vol. 12, no. 2, pp. 56–65, Apr. 2005.

[Hän20a] P. Händel, O. T. Demir, E. Björnson, and D. Rönnow, "Impact of backward crosstalk in 2×2 MIMO transmitters on NMSE and spectral efficiency", *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4277–4292, July 2020.

[Hän20b] P. Händel, and D. Rönnow, "MIMO and massive MIMO transmitter crosstalk", *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1882–1893, Mar. 2020.

[Hau17] K. Hausmair, S. Gustafsson, C. Sánchez-Pérez, Per N. Landin, U. Gustavsson, T. Eriksson, and C. Fager, "Prediction of nonlinear distortion in wideband active antenna arrays", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4550–4563, Nov. 2017.

[Hau18] K. Hausmair, P. N. Landin, U. Gustavsson, C. Fager, and T. Eriksson, "Digital predistortion for multi-antenna transmitters affected by antenna crosstalk", *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1524–1535, Mar. 2018.

[Hay09] S. S. Haykin, *Neural Networks and Learning Machines*, Pearson Education, Upper Saddle River, NJ, 3$^{\underline{\text{rd}}}$ ed., 2009.

[Hil00] G.R. Hill, M. Faulkner, and J. Singh, "Reducing the peak-to-average power ratio in OFDM by cyclically shifting partial transmit sequences", *Electron. Lett.*, vol. 36, no. 6, pp. 560, Mar. 2000.

[Hot33] H. Hotelling, "Analysis of a complex of statistical variables into principal components.", *J. Educ. Psychol.*, vol. 24, pp. 498–520, 1933.

[Hua12] G. Huang, S. Song, and C. Wu, "Orthogonal least squares algorithm for training cascade neural networks", *IEEE Trans. Circuits Syst. I, Reg. Paperss*, vol. 59, no. 11, pp. 2629–2637, Nov. 2012.

[Hus93] D. R. Hush, and B.G. Horne, "Progress in supervised neural networks", *IEEE Signal Process. Mag.*, vol. 10, no. 1, pp. 8–39, Jan. 1993.

[Isa05] M. Isaksson, D. Wisell, and D. Rönnow, "Wide-band dynamic modeling of power amplifiers using radial-basis function neural networks", *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 11, pp. 3422–3428, Nov. 2005.

[Jar18a] P. Jaraut, M. Rawat, and F. M. Ghannouchi, "Composite neural network digital predistortion model for joint mitigation of crosstalk, I/Q imbalance, nonlinearity in MIMO transmitters", *IEEE Trans. Microw. Theory Techn.*, pp. 1–10, Nov. 2018.

[Jar18b] P. Jaraut, M. Rawat, and F. M. Ghannouchi, "Harmonically related concurrent triband behavioral modeling and digital predistortion", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, pp. 1–5, Oct. 2018.

[Jia04] T. Jiang, and G. Zhu, "OFDM peak-to-average power ratio reduction by complement block coding scheme and its modified version", *IEEE 60th Veh. Technol. Conf., 2004*, IEEE, 2004.

[Jia05] T. Jiang, and G. Zhu, "Complement block coding for reduction in peak-to-average power ratio of OFDM signals", *IEEE Commun. Mag.*, vol. 43, no. 9, pp. S17–S22, Sept. 2005.

[Jol02] I.T. Jolliffe, *Principal Component Analysis*, Springer Series in Statistics, Springer, 2002.

[Jon93] S. de Jong, "SIMPLS: An alternative approach to partial least squares regression", *Chemom. Intell. Lab. Syst.*, vol. 18, no. 3, pp. 251–263, Mar. 1993.

[Jue17] P. Jueschke, and G. Fischer, "Machine learning using neural networks in digital signal processing for RF transceivers", *in Proc. AFRICON Conf.*, IEEE, Sept. 2017.

[Kah52] L. R. Kahn, "Single-Sideband Transmission by Envelope Elimination and Restoration", *in Proc. Inst. Radio Eng.*, vol. 40, no. 7, pp. 803–806, July 1952.

[Kar06] O. Karray, and C. De Silva, "Soft computing and intelligent systems design, theory, tools and applications", *IEEE Trans. Neural Netw.*, vol. 17, no. 3, pp. 825–825, May 2006.

[Kat16] A. Katz, J. Wood, and D. Chokola, "The evolution of PA linearization: From classic feedforward and feedback through analog and digital predistortion", *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 32–40, Feb. 2016.

[Kha14] A. Khattab, and M. A. Bayoumi, "The challenges towards energy-efficient cognitive radio networking", *IEEE 12th Int. New Circ. Syst. Conf.*, pp. 221–224, 2014.

[Kha17] Z. A. Khan, E. Zenteno, P. Händel, and M. Isaksson, "Digital predistortion for joint mitigation of I/Q imbalance and MIMO power amplifier distortion", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 1, pp. 322–333, Jan. 2017.

[Kim01] J. Kim, and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory", *Electron. Lett.*, vol. 37, pp. 1417–1418, Nov. 2001.

[Kim06] W.-J. Kim, K.-J. Cho, S. P. Stapleton, and J.-H. Kim, "Piecewise pre-equalized linearization of the wireless transmitter with a Doherty amplifier", *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 9, pp. 3469–3478, Sept. 2006.

[Kim07a] W.-J. Kim, K.-J. Cho, S. P. Stapleton, and J.-H. Kim, "Doherty feed-forward amplifier performance using a novel crest factor reduction technique", *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 1, pp. 82–84, Jan. 2007.

[Kim07b] W.-J. Kim, K.-J. Cho, S. P. Stapleton, and J.-H. Kim, "An efficient crest factor reduction technique for wideband applications", *Analog Integr. Circ. Signal Process.*, vol. 51, no. 1, pp. 19–26, Apr. 2007.

[Kra17] O. Kramer, *Genetic Algorithms*, pp. 11–19, Springer International Publishing, Cham, 2017.

[Kra20] J. Kral, T. Gotthans, R. Marsalek, M. Harvanek, and M. Rupp, "On feedback sample selection methods allowing lightweight digital predistorter adaptation", *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 6, pp. 1976–1988, June 2020.

[Kro03] B.S. Krongold, and D.L. Jones, "PAR reduction in OFDM via active constellation extension", *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 258–268, Sept. 2003.

[Lan15] Per N. Landin, Sebastian Gustafsson, Christian Fager, and Thomas Eriksson, "Weblab: A web-based setup for PA digital predistortion and characterization [application notes]", *IEEE Microw. Mag.*, vol. 16, no. 1, pp. 138–140, Feb. 2015.

[LB14] D. López-Bueno, P. L. Gilabert, G. Montoro, and N. Bartzoudis, "Peak cancellation and digital predistortion of high-order QAM wideband signals for next generation wireless backhaul equipment", *in Proc. Int. Workshop Integr. Nonlinear Microw. Millimetre-w. Circ.*, IEEE, apr 2014.

[LB16] D. López-Bueno, T. Wang, P. L. Gilabert, and G. Montoro, "Amping up, saving power: Digital predistortion linearization strategies for power amplifiers under wideband 4g5g burst-like waveform operation", *IEEE Microw. Mag.*, vol. 17, no. 1, pp. 79–87, Jan. 2016.

[LB18] D. López-Bueno, Q. A. Pham, G. Montoro, and P. L. Gilabert, "Independent digital predistortion parameters estimation using adaptive principal component analysis", *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5771–5779, Dec. 2018.

[LB20] D. López-Bueno, N. Bartzoudis, O. Font-Bach, M. Caus, P. Gilabert, and G. Montoro, "Technologies for emergency rollout of broadband public protection and disaster relief (BB-PPDR) communications in humanitarian crisis zones", *Information and Communication Technologies for Humanitarian Services*, pp. 215–246, Institution of Engineering and Technology, Aug. 2020.

[LB21] D. López-Bueno, P. L. Gilabert, and G. Montoro, "Dataset reduction for neural network based digital predistorters under strong nonlinearities", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, IEEE, Jan. 2021.

[LB22a] D. López-Bueno, "SHAPER brochure", CTTC Adaptive Processing Technologies (ADAPT) Research Unit site. [Online]. Available: https://www.cttc.cat/wp-content/uploads/2022/02/Extended_Description_ADAPT-2.pdf (accessed May, 2023), Jan. 2022.

[LB22b] D. López-Bueno, G. Montoro, and P. L. Gilabert, "Training data selection and dimensionality reduction for polynomial and artificial neural network mimo adaptive digital predistortion", *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4940–4954, 2022.

[Li21] W. Li, E. Guillena, G. Montoro, and P. L. Gilabert, "FPGA implementation of memory based digital predistorters with high-level synthesis", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, pp. 1–4, jan. 2021.

[Li22] W. Li, N. Bartzoudis, J. Rubio Fernández, D. López-Bueno, G. Montoro, and P. L. Gilabert, "FPGA implementation of a linearization system for wideband envelope tracking power amplifiers", *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 4, pp. 1781–1792, Nov. 2022.

[Li23] W. Li, G. Montoro, and P. L. Gilabert, "Digital linearization of wideband envelope tracking power amplifiers for mobile terminals", *IEEE Trans. Microw. Theory Techn.*, vol. 71, no. 1, pp. 48–58, Jan. 2023.

[Mar15] M. A. Marotta, N. Kaminski, I. Gómez-Miguélez, L. Z. Granville, J. Rochol, L. DaSilva, and C. B. Both, "Resource sharing in heterogeneous cloud radio access networks", *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 74–82, June 2015.

[Mio18] M. Miozzo, N. Bartzoudis, M. Requena, O. Font-Bach, P. Harbanau, D. López-Bueno, M. Payaró, and J. Mangues, "SDR and NFV extensions in the ns-3 LTE module for 5G rapid prototyping", *2018 IEEE Wireless Commun. and Networking Conference (WCNC)*, pp. 1–6, 2018.

[Mka11] F. Mkadem, and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion", *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 4, pp. 913–923, Apr. 2011.

[Möl10] B. G. Mölleryd, J. Markendahl, J. Werding, and Ö. Mäkitalo, "Decoupling of revenues and traffic - is there a revenue gap for mobile broadband?", *9th Conf. Telecommun. Media Internet*, pp. 1–7, 2010.

[Mol17] 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 Trans. Microw. Theory Techn.*, vol. 65, no. 4, pp. 1381–1393, Apr. 2017.

[Mon07] G. Montoro, P. L. Gilabert, E. Bertran, A. Cesari, and D. D. Silveira, "A New Digital Predictive Predistorter for Behavioral Power Amplifier Linearization", *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 6, pp. 448–450, June 2007.

[Mor06] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers", *IEEE Trans. on Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.

[Mra12] N. Mrabet, I. Mohammad, F. Mkadem, C. Rebai, and S. Boumaiza, "Optimized hardware for polynomial digital predistortion system implementation", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, pp. 83–84, IEEE, Jan. 2012.

[MTA19] Australian Association Mobile Telecommunications (AMTA), GSMA, and Mobile & Wireless Forum (MWF), "5G and EMF explained", EMF explained series. [Online]. Available: https://www.emfexplained.info/ (accessed May, 2023), Aug. 2019.

[Mül97] S.H. Müller, and J.B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences", *Electron. Lett.*, vol. 33, no. 5, pp. 368, Feb. 1997.

[Mur06] S.D. Muruganathan, and A.B. Sesay, "A QRD-RLS-based predistortion scheme for high-power amplifier linearization", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 10, pp. 1108–1112, Oct. 2006.

[NXP13] NXP, "High performance RF for wireless infrastructure", NXP Semiconductor N.V. Unleash RF series white paper. [Online]. Available: http://www.northstardatasheet.com/datasheet/75017347.pdf (accessed May, 2023), Jan. 2013.

[Paa08] H. Paaso, and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures", *IEEE Int. Symp. Wireless Commun. Syst.*, pp. 309–313, 2008.

[Pen16a] J. Peng, S. He, B. Wang, Z. Dai, and J. Pang, "Digital predistortion for power amplifier based on sparse Bayesian learning", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 63, no. 9, pp. 828–832, Sept. 2016.

[Pen16b] R. Pengelly, C. Fager, and M. Ozen, "Doherty's legacy: A history of the doherty power amplifier from 1936 to the present day", *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 41–58, Feb. 2016.

[Pha18a] Q. A. Pham, D. López-Bueno, G. Montoro, and P. L. Gilabert, "Adaptive principal component analysis for online reduced order parameter extraction in PA behavioral modeling and DPD linearization", *in IEEE MTT-S Int. Microw. Symp. Dig.*, pp. 160–163, Jun. 2018.

[Pha18b] Q. A. Pham, D. López-Bueno, T. Wang, G. Montoro, and P. L. Gilabert, "Multidimensional LUT-based digital predistorter for concurrent dual-band envelope tracking power amplifier linearization", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, pp. 47–50, Jan. 2018.

[Pha18c] Q. A. Pham, D. López-Bueno, T. Wang, G. Montoro, and P. L. Gilabert, "Partial least squares identification of multi look-up table digital predistorters for concurrent dual-band envelope tracking power amplifiers", *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5143–5150, Dec. 2018.

[Pha19a] Q. A. Pham, D. López-Bueno, G. Montoro, and P. L. Gilabert, "Dynamic selection and update of digital predistorter coefficients for power amplifier linearization", *in Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl.*, pp. 1–4, Jan. 2019.

[Pha19b] Q. A. Pham, G. Montoro, D. López-Bueno, and P. L. Gilabert, "Dynamic selection and estimation of the digital predistorter parameters for power amplifier linearization", *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 3996–4004, Oct. 2019.

[Pla86] D. C. Plaut, S. J. Nowlan, and G. E. Hinton, "Experiments on learning by back propagation", *Technical Report CMU-CS-86-126*, Carnegie-Mellon University, 1986.

[Poi20] V. Poirot, M. Ericson, M. Nordberg, and K. Andersson, "Energy efficient multi-connectivity algorithms for ultra-dense 5G networks", *Wireless Networks*, vol. 26, no. 3, pp. 2207–2222, Apr. 2020.

[Pop17] Z. Popovic, "Amping up the PA for 5G: Efficient GaN power amplifiers with dynamic supplies", *IEEE Microw. Mag.*, vol. 18, no. 3, pp. 137–149, May 2017.

[Pot99] N. Pothecary, *Feedforward Linear Power Amplifiers*, Artech House Inc., 685 Canton St., Norwood, MA, 1999.

[Qua22] R. Quaglia, J. Pang, S. C. Cripps, and A. Zhu, "Load-modulated balanced amplifier: From first invention to recent development", *IEEE Microw. Mag.*, vol. 23, no. 12, pp. 60–70, Dec. 2022.

[Raa01] F. H. Raab, "Class-E, class-C, and class-F power amplifiers based upon a finite number of harmonics", *IEEE Trans. Microw. Theory Techn.*, vol. 49, no. 8, pp. 1462–1468, Aug. 2001.

[Rah13] Y. Rahmatallah, and S. Mohan, "Peak-to-average power ratio reduction in ofdm systems: A survey and taxonomy", *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1567–1592, Fourth Quarter 2013.

[Raw10] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks", *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 1, pp. 95–104, Jan. 2010.

[Rho03] J. D. Rhodes, "Output universality in maximum efficiency linear power amplifiers", *Int. J. Circuit Theory Appl.*, vol. 31, no. 4, pp. 385–405, 2003.

[Rit20] H. Ritchie, "Climate change and flying: what share of global co2 emissions come from aviation?", Our World in Data. [Online]. Available: https://ourworldindata.org/co2-emissions-from-aviation (accessed May, 2023), Oct. 2020.

[Riz11] in Proc. XXVI Simp. Nac. Unión Cient. Int. Radio, 2011.

[Rob13] P. Roblin, C. Quindroit, N. Naraharisetti, S. Gheitanchi, and M. Fitton, "Concurrent linearization: The state of the art for modeling and linearization of multiband power amplifiers", *IEEE Microw. Mag.*, vol. 14, no. 7, pp. 75–91, Nov. 2013.

[Ros58] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain.", *Psychol. Rev.*, vol. 65, no. 6, pp. 386–408, 1958.

[RT15] J. Reina-Tosina, M. Allegue-Martínez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis", *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 745–753, Feb. 2015.

[Rum86] D. E. Rumelhart, and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (2 Volume Set) (Vol.1)*, Mit Pr, 1986.

[S.01] Karl Pearson F. R. S., "LIII. On lines and planes of closest fit to systems of points in space", *Philos. Mag. Ser. 1*, vol. 2, pp. 559–572, 1901.

[Saf11] D. Saffar, N. Boulejfen, F. M. Ghannouchi, A. Gharsallah, and M. Helaoui, "Behavioral modeling of MIMO nonlinear systems with multivariable polynomials", *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 11, pp. 2994–3003, Nov. 2011.

[Sal83] A. A. M. Saleh, and J. Salz, "Adaptive linearization of power amplifiers in digital radio systems", *Bell Syst. Techn. J.*, vol. 62, no. 4, pp. 1019–1033, Apr. 1983.

[Sau04] A. Saul, "Peak reduction for OFDM by shaping the clipping noise", *IEEE 60th Vehic. Technol. Conf.*, IEEE, 2004.

[Sch09] D. Scheurs, M. O'Droma, A. A. Goacher, and M. Gadringer (eds.), *RF Power Amplifier Behavioural Modeling*, (The Cambridge RF and Microwave Engineering Series). Cambridge: Cambridge University Press., 2009.

[Sha07] A. Sharma, and K. K. Paliwal, "Fast principal component analysis using fixed-point algorithm", *Pattern Recognit. Lett.*, vol. 28, no. 10, pp. 1151–1155, 2007.

[She16] D. J. Shepphard, J. Powell, and S. C. Cripps, "An efficient broadband reconfigurable power amplifier using active load modulation", *IEEE Microw. Wireless Compon. Lett.*, vol. 26, no. 6, pp. 443–445, June 2016.

[Shi16] V. Shilimkar, and G. Wang, "2015 ims student design competitions", *IEEE Microw. Mag.*, vol. 17, no. 1, pp. 40–41, Jan. 2016.

[Sow04] T. Sowlati, D. Rozenblit, E. MacCarthy, M. Damgaard, R. Pullela, D. Koh, and D. Ripley, "Quad-band gsm/gprs/edge polar loop transmitter", *IEEE Int. Solid-State Circ. Conf. (IEEE Cat. No.04CH37519)*, pp. 186–521 Vol.1, 2004.

[Sto78] M. Stone, "Cross-validation: a review", *Ser. Statistics*, vol. 9, no. 1, pp. 127–139, Jan. 1978.

[Tay13] A. Taylor, "There was something different about the vatican crowd in 2005", Business Insider. [Online]. Available: http://www.businessinsider.com/vatican-square-2005-and-2013-2013-3. (accessed May, 2023), Mar. 2013.

[Teh12] A. S. Tehrani, T. Eriksson, and C. Fager, "Modeling of long term memory effects in RF power amplifiers with dynamic parameters", *2012 IEEE/MTT-S Int. Microw. Symp. Dig.*, pp. 1–3, 2012.

[Tho16] A. J. Thomas, S. D. Walters, S. M. Gheytassi, R. E. Morgan, and M. Petridis, "On the optimal node ratio between hidden layers: A probabilistic study", *Int. J. Mach. Learn. Comput.*, vol. 6, no. 5, pp. 241–247, oct 2016.

[TM99] J. Tellado-Mourelo, *Peak to Average Power Ratio Reduction for Multicarrier Modulation*, PhD Thesis, Department of Electrical Engineering, University of Stanford, 1999.

[Tre97] L. N. Trefethen, and D. Bau, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[TSIE14] European Telecommunications Standards Institute (ETSI), *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Conformance Testing*, 3GPP TS 36.141 V12.5.0 R12 Technical Specification, Oct. 2014.

[Vaa05] O. Vaananen, J. Vankka, and K. Halonen, "Simple algorithm for peak windowing and its application in GSM, EDGE and WCDMA systems", *IEEE Proceedings - Communications*, vol. 152, no. 3, pp. 357, 2005.

[Wal12] J. L. B. Walker, *Handbook of RF and Microwave Power Amplifiers*, The Cambridge RF and Microwave Engineering Series, Cambridge: Cambridge University Press, 2012.

[Wan15] T. Wang, P. L. Gilabert, and G. Montoro, "Under-sampling effects and computational cost reduction in rf power amplifier behavioral modeling", *10th Eur. Microw. Int. Circ. Conf.*, pp. 57–60, 2015.

[Wan17a] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low computational complexity digital predistortion based on direct learning with covariance matrix", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4274–4284, Nov. 2017.

[Wan17b] Z. Wang, L. Guan, and R. Farrell, "Undersampling observation-based compact digital predistortion for single-chain multiband and wideband direct-to-RF transmitter", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 12, pp. 5274–5283, Dec. 2017.

[Wan19] D. Wang, M. Aziz, M. Helaoui, and F. M. Ghannouchi, "Augmented real-valued time-delay neural network for compensation of distortions and impairments in wireless transmitters", *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 1, pp. 242–254, Jan. 2019.

[Wan21a] T. Wang, and P. L. Gilabert, "Mesh-selecting for computational efficient PA behavioral modeling and DPD linearization", *IEEE Microw. Wireless Compon. Lett.*, vol. 31, no. 1, pp. 37–40, Jan. 2021.

[Wan21b] T. Wang, W. Li, R. Quaglia, and P. L. Gilabert, "Machine-learning assisted optimisation of free-parameters of a dual-input power amplifier for wideband applications", *Sensors*, vol. 21, no. 8, 2021.

[Wan21c] X. Wang, Y. Li, H. Yin, C. Yu, Z. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G multiuser MIMO transmitters using low-dimensional feature-based model generation", *IEEE Trans. Microw. Theory Techn.*, pp. 1–1, Mar. 2021.

[Wes19] E. Westberg, J. Staudinger, J. Annes, and V. Shilimkar, "5G Infrastructure RF solutions: Challenges and opportunities", *IEEE Microw. Mag.*, vol. 20, no. 12, pp. 51–58, Mar. 2019.

[Wis08] D. Wisell, J. Jalden, and P. Händel, "Behavioral power amplifier modeling using the lasso", *2008 IEEE Inst. and Meas. Tech. Conf.*, pp. 1864–1867, May 2008.

[Woo14] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*, Artech House Microwave Library, Artech House Inc., 685 Canton St., Norwood, MA, 2014.

[Woo17] J. Wood, "System-level design considerations for digital pre-distortion of wireless base station transmitters", *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1880–1890, May 2017.

[Yan21] G. Yang, F. Liu, H. Li, W. Qiao, C. Jiang, and L. Su, "Sample selection method for digital predistortion with incremental dimension of coefficients", *IEEE MTT-S Int. Wireless Symp.*, IEEE, May. 2021.

[Yu19] C. Yu, J. Jing, H. Shao, Z. H. Jiang, P. Yan, X.-W. Zhu, W. Hong, and A. Zhu, "Full-angle digital predistortion of 5G millimeter-wave massive MIMO transmitters", *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 7, pp. 2847–2860, July 2019.

[Zay10] R. Zayani, R. Bouallegue, and D. Roviras, "Crossover neural network predistorter for the compensation of crosstalk and nonlinearity in MIMO OFDM systems", *21st Annu. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, IEEE, 2010.

[Zen14] E. Zenteno, S. Amin, M. Isaksson, D. Rönnow, and P. Händel, "Combating the dimensionality of nonlinear MIMO amplifier predistortion by basis pursuit", *in Proc. Eur. Microw. Conf.*, IEEE, 2014.

[Zha97] Y. Zhang, and Y. Ma, "CGHA for principal component extraction in the complex domain", *IEEE Trans. Neural Net.*, vol. 8, no. 5, pp. 1031–1036, Sept. 1997.

[Zhu06] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of RF power amplifiers", *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[Zhu08] A. Zhu, P. J. Draxler, C. Hsia, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Digital predistortion for envelope-tracking power amplifiers using decomposed piecewise volterra series", *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 10, pp. 2237–2247, Oct. 2008.

[Zhu16] R. Zhu, "Gradient-based sampling: An adaptive importance sampling for least-squares", *in Proc. Int. Conf. Neural Inf. Process. Syst.*, NIPS'16, pp. 406–414, Curran Associates Inc., Red Hook, NY, USA, 2016.