Tampere University

PABLO PASCUAL CAMPO

Low-complexity Algorithms for Digital Predistortion and Digital Self-interference Cancellation

Tampere University Dissertations 532

Tampere University Dissertations 532

PABLO PASCUAL CAMPO

Low-complexity Algorithms for Digital Predistortion and Digital Self-interference Cancellation

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Information Technology and Communication Sciences
of Tampere University,
for public discussion in the Auditorium TB109
of the Tietotalo Building, Korkeakoulunkatu 1, Tampere,
on the 28th January 2022, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Information Technology and Communication Sciences
Finland

Responsible supervisor and Custos	Prof. Mikko Valkama Tampere University Finland	
Supervisors	Dr. Lauri Anttila Tampere University Finland	
Pre-examiners	Assoc. Prof. Pere Lluis Gilabert Pinal Universitat Politècnica de Catalunya Spain	Prof. Timo Erkki Rahkonen University of Oulu Finland
Opponent	Prof. Daniel Rönnow University of Gävle Sweden	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2021 author

Cover design: Roihu Inc.

ISBN 978-952-03-2235-9 (print)
ISBN 978-952-03-2236-6 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2236-6

PunaMusta Oy – Yliopistopaino
Joensuu 2021

To my parents, Pilar and Jordi.

PREFACE

THE research work presented within this thesis was carried out during the years 2018-2021, at the Faculty of Electrical Engineering, Tampere University, Tampere, Finland. This thesis work was financially supported by the Academy of Finland (under the projects #304147 "In-Band Full-Duplex Radio Technology: Realizing Next Generation Wireless Transmission", #301820 "Competitive Funding to Strengthen University Research Profiles", #319994 "Research Infrastructure for Future Wireless Communication Networks", #323461 "Hybrid amplitude/time-based transmitter for future wireless communications", and #332361 "Energy-efficient Radio Access: Methods and Optimization"), Tampere University Doctoral School, Finnish Funding Agency for Innovation (Tekes, under the projects "5G Transceivers for Base Stations and Mobile Devices (5G TRx)", and "TAKE-5"), Huawei Technologies Finland, European Union's Horizon 2020 research and innovation program (under grant agreement #732174 "ORCA, Extension #4") and the Nokia Foundation. I would like to acknowledge all of these projects for their financial contribution, which, in part, made this work possible.

First of all, I would like to express my most sincere gratitude to my thesis supervisor, Prof. Mikko Valkama, who has guided me throughout these years, providing exceptional advice, help, and showing great patience. His commitment and discipline have made me push and improve myself every single day, and for that I am truly grateful. I feel really lucky to also have worked with Dr. Lauri Anttila, who has spared no time in answering my endless inquiries and questions, and who has also offered many creative suggestions and ideas which have helped to improve the overall quality of my work. I also want to thank Dr. Dani Korpi, who not only greatly assisted me at the beginning of the research process, but has also been an example to follow during these years.

I am also grateful to my thesis pre-examiners, Assoc. Prof. Pere Lluis Gilabert

Pinal, and Prof. Timo Rahkonen, for their valuable comments and suggestions, which have helped me to polish my thesis work and thus improve its overall quality. Furthermore, I would also like to thank Prof. Daniel Rönnow, for agreeing to be the opponent of the thesis examination.

I wish also to thank my every-day coworkers and colleagues from Tampere University for providing a wonderful working environment, full of interesting discussions and ideas. In particular, I would like to thank M.Sc. Vesa Lampu and M.Sc. Selahattin Gökceli, for providing endless moments of fun and laughter, and for those stays in the cottage which I will never forget. In addition, I would like to thank M.Sc. Carlos Baquero, M.Sc. Alberto Brihuega, M.Sc. Duc Le, M.Sc. Matias Turunen, M.Sc. Sahan Liyanaarachchi, M.Sc. Nachiket Ayir, Dr. Guixian Xu, and Dr. Sener Dikmese, among others, for accompanying me on this journey.

There are also uncountable friends whose names are not included in this Preface, but whom I should also acknowledge. I thank all of my friends who remained in Spain, specially Miguel García and Guillermo Mateo. Your constant good vibes have been essential to keep my motivation always up, and to always remember that life is fun. I would also like to deeply thank Loreena McKennitt, for providing quality music which kept my head focused when I needed it most, and also Ghost, for always boosting my motivation, and for making me forget my worries at all times. I should not forget to mention my gratitude towards the city of Cape Town, which has become my preferred holiday destination in the last years, always offering a quiet and ideal place to relax and disconnect from the world. Furthermore, I would like to also thank María Chicote, for helping me uncountable times, for lifting me up when I was down, and for also celebrating with me at my successes.

Last, and most importantly, I would like to give my most sincere and deepest gratitude to my parents, Pilar and Jordi. You have guided me throughout my whole life, always supporting me and pointing me to the right direction, no matter what obstacles life may have presented. The endless effort you have made supporting me in my best and worst moments has never gone unnoticed. I strongly believe that I could not have written this dissertation without your help, hence, I ask you sincerely to consider this work as also your own. As a child, I may have sometimes thought it to be too much, but without the strong discipline, self-sacrifice, and willpower you have taught me I could not have aspired to where I am now. Thank you, mom,

for your endless advice, and for making me the man I am today. Thank you, dad, for our endless discussions about everything, and for your numerous “sermones”, which have undoubtedly left a gigantic mark on my character. I am proud to have you as my parents.

Tampere, November 2021
Pablo Pascual Campo

ABSTRACT

JUDGING from the massive impact wireless communications made during the past decades, there is little doubt that it will still be one of the major foundations of contemporary society also in the future. In the context of modern radio networks, two main goals are currently pursued. First, a more efficient usage of the extremely scarce radio spectrum, with the aim of really exploiting the usable frequency ranges. Second, an improved cost-effective trade-off of the physical transmitters, with the aim of minimizing the size and cost of the involved circuitry, and also to facilitate more power-efficient solutions which will reduce the overall expenses of wireless networks and also aid in the reduction of carbon emissions.

In order to improve the system spectral efficiency and transmitter energy efficiency, digital predistortion (DPD) and digital self-interference (SI) cancellation techniques have been widely proposed in the literature, however, research in associated low-complexity solutions is more scarce. With reduced-complexity algorithms being a new trend in 5G and beyond networks, it becomes crucial to explore solutions which minimize the computational complexities that are involved. This claim is motivated by the following three trends. First, the utilized signal bandwidths are becoming wider and wider in emerging networks, thus also increasing DPD processing rates. Second, it is a clear new preference to have a smaller base station (BS) configuration, thus also the available power budget dedicated for front-end digital processing techniques is reduced. Third, many solutions require fast adaptation of the model coefficients, such as the SI channel in in-band full duplex (IBFD), or the DPD coefficient estimation in mmW beam-steered antenna array systems.

To this end, this thesis firstly concentrates on providing several low-complexity modeling solutions which are applicable for both DPD and IBFD applications. In particular, four cascaded structures are proposed and combined with a complex injection-based spline-interpolation scheme, which is able to provide accurate modeling of the PA-induced nonlinearity. Additional memory effects are modeled by elemen-

tary linear time-invariant adaptive filters. The cascaded structures also incorporate efficient gradient-descent-based algorithms to adaptively update the model parameters. Furthermore, a look-up table-based memory polynomial (MP) structure is proposed, which alternatively models the memory through a MP-like parallel branched structure. This approach provides richer modeling capabilities, and it does not require any physical knowledge of the system under study. In the context of DPD, three reduced-complexity signed closed-loop techniques are devised to reduce the complexity of the learning path. Additionally, different inverse covariance matrix estimation methods are presented in the context of a self-orthogonalized learning system, which reduce its computational complexity even further.

Secondly, this thesis presents exhaustive and comprehensive RF verification and validation, which, together with detailed complexity analyses, allow assessment of the performance-complexity trade-offs of the proposed methods. In the context of DPD, the proposed solutions are tested with two different frequency range 1 (FR-1) and frequency range 2 (FR-2) RF measurement environments, considering also different off-the-shelf PA systems. In all cases, it is shown how the proposed methods enhance the energy efficiency of the transmitter, while keeping the unwanted in-band and out-of-band distortion under the levels specified in the latest 3GPP 5G NR Release 15. In the context of IBFD, corresponding evaluations are carried out with two actual real-life IBFD prototypes, showing that the proposed techniques achieve a similar cancellation to other state-of-the-art techniques, regardless of the drastic complexity reductions.

Altogether, the strong mathematical foundations of the developed solutions, together with the obtained results, show that more efficient yet reliable transmitters can be achieved through the proposed DPD methods, allowing for the minimization of cost and size of the circuitry involved, and improving the energy efficiency of the transmitter. Additionally, this thesis verifies the commercial feasibility of IBFD, considering that the SI is suppressed. SI suppression is achieved by the proposed methods, which bring IBFD one step closer to commercially feasible implementations.

TABLE OF CONTENTS

1 Introduction 1
1.1 Thesis Motivation and Scope 1
1.2 Thesis Objectives 5
1.3 Thesis Contributions and Structure 6
1.4 Author's Contributions to the Publications 7
1.5 Nomenclature 9

2 Background and State of the Art 11
2.1 Trends and Challenges in Modern Wireless Communications 11
2.1.1 PA Nonlinear Distortion Quantification 12
2.1.2 SI Quantification in IBFD Applications 14
2.2 Digital Predistortion 16
2.2.1 Basics of DPD 16
2.2.2 State-of-the-art in DPD and Linearization Techniques 22
2.3 Digital SI Cancellation for IBFD Transceivers 26
2.3.1 Basics of Digital SI Cancellation 26
2.3.2 State-of-the-art in Digital SI Cancellation 29

3 Spline-based Cascaded Models 33
3.1 Review of B-spline Interpolation Theory 33
3.1.1 Real-valued B-spline Interpolation 33
3.1.2 Complex-valued B-spline Interpolation 36
3.1.3 Injection-based Complex-valued B-spline Interpolation 38
3.2 Spline-based Hammerstein Model 40

3.2.1 Learning Equations 41
3.2.2 Complexity Analysis 43
3.3 Spline-based Wiener Model 44
3.3.1 Learning Equations 45
3.3.2 Complexity Analysis 46
3.4 Spline-based Wiener-Hammerstein Model 46
3.4.1 Learning Equations 48
3.4.2 Complexity Analysis 49
3.5 Spline-based Hammerstein-Wiener Model 50
3.5.1 Learning Equations 50
3.5.2 Complexity Analysis 52

4 Linear-in-parameters Models 53
4.1 Spline-based MP-LUT Model 53
4.1.1 SPMP Processing Scheme 55
4.1.2 Learning Equations 55
4.1.3 Complexity Analysis 56
4.2 MP-LUT Model 57
4.2.1 MP-LUT Processing Scheme 58
4.2.2 Classical Closed-loop Learning Algorithms 59
4.3 Closed-loop Signed Learning Algorithms 60
4.3.1 Sign Algorithm 61
4.3.2 Sign Regressor Algorithm 62
4.3.3 Sign-sign Algorithm 63
4.3.4 Complexity Analysis 64
4.4 ICM Estimation Methods in SO Learning Algorithm 66
4.4.1 Estimating the ICM with the Autocorrelation Function . . . 68
4.4.2 Approximating the ICM with the Bussgang Theorem 70
4.4.3 Reducing the Dimensionality of the ICM 74
4.4.4 Frequency Shifting the ICM 75

4.4.5 Recursively Estimating the ICM 77
4.4.6 Complexity Analysis 79

5 Experimental Verification 81
5.1 RF Verification in Digital Predistortion 81
5.1.1 Deployed RF Measurement Setups 81
5.1.2 RF Verification of Spline-based Cascaded Models 85
5.1.3 RF Verification of Spline-based Linear-in-parameters Solutions 88
5.1.4 RF Verification of the ICM Estimation Methods 93
5.2 RF Verification in Digital Self-interference Cancellation 94
5.2.1 RF Verification of Spline-based Cascaded Models 94
5.2.2 RF Verification of Spline-based MP-LUT Models 100

6 Conclusion 103
6.1 Main Findings 103
6.2 Future Work 105

Bibliography 109

Appendix AM/AM Response of the Tested PA Systems 141

LIST OF PUBLICATIONS

P1 P. Pascual Campo, A. Brihuega, L. Anttila, M. Turunen, D. Korpi, M. Allén, and M. Valkama, "Gradient-adaptive spline-interpolated LUT methods for low-complexity digital predistortion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 1, pp. 336–349, Jan. 2021. DOI: 10.1109/TCSI.2020.3034825.

P2 P. Pascual Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allén, Y. Guo, and M. Valkama, "Closed-loop sign algorithms for low-complexity digital predistortion: Methods and performance," *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 1, pp. 1048–1062, Jan. 2021. DOI: 10.1109/TMTT.2020.3038316.

P3 P. Pascual Campo, L. Anttila, D. Korpi, and M. Valkama, "Cascaded spline-based models for complex nonlinear systems: Methods and applications," *IEEE Transactions on Signal Processing*, vol. 69, pp. 370–384, Feb. 2021. DOI: 10.1109/TSP.2020.3046355.

P4 P. Pascual Campo, L. Anttila, V. Lampu, Y. Guo, N. Wang, and M. Valkama, "Inverse covariance matrix estimation for low-complexity closed-loop DPD systems: Methods and performance," *IEEE Transactions on Microwave Theory and Techniques*, 2021. DOI: 10.1109/TMTT.2021.3124226.

P5 P. Pascual Campo, D. Korpi, L. Anttila, and M. Valkama, "Nonlinear digital cancellation in full-duplex devices using spline-based Hammerstein model," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018, pp. 1–7. DOI: 10.1109/GLOCOMW.2018.8644362.

P6 P. Pascual Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allén, and M. Valkama, "Closed-loop sign algorithms for low-complexity digital predistor-

tion," in *2020 IEEE/MTT-S International Microwave Symposium (IMS)*, Jun. 2020, pp. 841–844. DOI: 10.1109/IMS30576.2020.9223904.

P7 P. Pascual Campo, L. Anttila, D. Korpi, and M. Valkama, "Adaptive cancellation of nonlinear self-interference in wireless full-duplex: Cascaded spline-interpolated methods," in *2020 54th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA*, Dec. 2020, pp. 1265–1271. DOI: 10.1109/IEEECONF51394.2020.9443406.

P8 P. Pascual Campo, L. Anttila, V. Lampu, Y. Guo, N. Wang, and M. Valkama, "Efficient inverse covariance matrix estimation for low-complexity closed-loop DPD systems," in *2021 IEEE MTT-S International Wireless Symposium (IWS)*, Apr. 2021, pp. 1–3. DOI: 10.1109/IWS52775.2021.9499480.

All of the articles presented above are published. In addition to these publications, there are several related works, specifically the publications contained in [21], [68], [98], [146], [155], [200], [244], [245], which have been also published thanks to the research work carried out during this dissertation process.

ABBREVIATIONS

3GPP	3rd generation partnership project
4G	Fourth generation
5G	Fifth generation
ACLR	Adjacent channel leakage ratio
ADC	Analog-to-digital converter
AM/AM	Amplitude-to-amplitude modulation
AM/PM	Amplitude-to-phase modulation
ANN	Artificial neural networks
ASIC	Application-specific integrated circuit
AWG	Arbitrary waveform generator
BF	Basis function
BLMS	Block least mean squares
BPF	Band-pass filter
BS	Base station
CA	Carrier aggregation
CCDF	Complementary cumulative distribution function
CDF	Complementary distribution function
CM	Covariance matrix
DAC	Digital-to-analog converter
DDR	Dynamic deviation reduction
DL	Downlink

DLA	Direct learning architecture
DPD	Digital predistortion
DSIC	Digital self-interference cancellation
DUT	Device under test
EBD	Electrical-balance duplexer
EIRP	Effective isotropic radiated power
EMP	Envelope memory polynomial
ET PA	Envelope tracking power amplifier
EVM	Error vector magnitude
FFT	Fast Fourier transform
FIR	Finite impulse response
FLOP	Floating point operation
FPGA	Field programmable gate array
FR-1	Frequency range 1
FR-2	Frequency range 2
GMP	Generalized memory polynomial
GN	Gauss-Newton
GPU	Graphics processor unit
I/Q	In-phase/quadrature
IBFD	In-band full-duplex
ICM	Inverse covariance matrix
IEEE	Institute of Electrical and Electronics Engineers
IF	Intermediate frequency
ILA	Indirect learning architecture
LDMOS	Laterally-diffused metal-oxide semiconductor
LMS	Least mean squares
LNA	Low noise amplifier

LO	Local oscillator
LPF	Low-pass filter
LS	Least squares
LTE	Long-term evolution
LTE-A	Long-term evolution-advanced
LTI	Linear time-invariant
LUT	Look-up table
M2M	Machine to machine
MIMO	Multiple-input multiple-output
mmW	Millimeter wave
MP	Memory polynomial
MP-LUT	Look-up table-based memory polynomial
MRAS	Model reference adaptive system
MTC	Machine type communication
NARMA	Nonlinear auto-regressive moving average model
NMSE	Normalized mean square error
NR	New radio
OFDM	Orthogonal frequency-division multiplexing
OoB	Out-of-band
ORX	Observation receiver
OTA	Over-the-air
PA	Power amplifier
PAPR	Peak-to-average power ratio
PDF	Probability distribution function
PH	Parallel Hammerstein
PIM	Passive intermodulation
PSD	Power spectral density

RB	Resource block
RF	Radio frequency
RLS	Recursive least squares
RX	Receiver
SCS	Subcarrier spacing
SI	Self-interference
SINR	Signal-to-interference plus noise ratio
SISO	Single-input single-output
SO	Self-orthogonalized
SPH	Spline-based Hammerstein
SPHW	Spline-based Hammerstein-Wiener
SPMP	Spline-based memory polynomial model
SPW	Spline-based Wiener
SPWH	Spline-based Wiener-Hammerstein
SRA	Sign regressor algorithm
STAR	Simultaneous transmission and reception
TRP ACLR	Total radiated power adjacent channel leakage ratio
TX	Transmitter
UE	User equipment
UL	Uplink
URLLC	Ultra reliable low-latency communications
VST	Vector signal transceiver

SYMBOLS

$\mathbf{0}_N$	Column vector consisting of N zeros
$\mathbf{1}_N$	Column vector consisting of N ones
$\mathbf{c}_n$	Complex control points of the spline interpolation scheme
csgn$(\cdot)$	Modified complex-valued sign function
C	Number of model coefficients of the input BF matrix $\mathbf{\Omega}$
C^{SP}	Number of control points in a spline-interpolated scheme
C_{IBFD}	Channel capacity in an IBFD transceiver
$\mathbf{C}_P$	B-spline basis matrix
$d[n]$	Desired signal in a digital signal model after the ADC
$d(t)$	Observed signal in an IBFD transceiver
$e[n]$	Error signal in a baseband architecture
$\mathbf{e}_k$	Block error signal in the kth learning iteration
f	Frequency variable
$f_{p,m_1,m_2,\cdots,m_p}$	Volterra kernels in the Volterra series
f_s	Sampling frequency
$f(\cdot)$	Frequency response of a nonlinear system
$g[k]$	kth linear FIR filter coefficient of $\mathbf{g}_n$
$\mathbf{g}_n$	Linear FIR filter coefficients used in the cascaded models
$G(\lvert x[n]\rvert)$	Complex nonlinear gain of a device, with input signal $x[n]$
$G_{\mathrm{I}}(\lvert x[n]\rvert)$	Real-valued gain of the I branch of $x[n]$
$G_{\mathrm{Q}}(\lvert x[n]\rvert)$	Real-valued gain of the Q branch of $x[n]$

$h_{\mathrm{SI}}(t)$	SI channel in an IBFD transceiver
$H(z^{-1})$	Transfer function of an LTI system
i_n	Index value in the spline interpolation scheme at time instant n
$J(x)$	Cost function dependent on x
k	Iteration index in a closed-loop learning rule
K	Spline interpolation order
K^{it}	Iteration sample size
$l[n]$	Second intermediate signal in a cascaded system
l^{ICM}	Number of rows and columns to be removed from $\tilde{\mathbf{Q}}^{-1}$
$\mathbf{l}_n$	Signal regression of $l[n]$
$\mathbf{L}_n$	Diagonal matrix containing the signal regression of $l[n]$
M	Total number of memory taps
M_{pre}	Number of pre-cursor taps
M_{post}	Number of post-cursor taps
M_{g}	Number of memory taps in the FIR filter $\mathbf{g}_n$
M_{w}	Number of memory taps in the FIR filter $\mathbf{w}_n$
M_{H}	Total number of memory taps in a Hammerstein model
M_{L}	Total number of memory taps in a nested LUT model
M_{MP}	Total number of memory taps in a MP model
M_{V}	Total number of memory taps in a Volterra model
M_{W}	Total number of memory taps in a Wiener model
n	Sample index
N	Parameter estimation sample size
N^{SP}	Number of regions in the spline-interpolated scheme
N_i^P	Pth-degree spline basis function characterized over a region i
P	Spline interpolation order
$\mathbf{P}$	Weight matrix in the ICM frequency shifting method

P_{H}	Polynomial order in a Hammerstein model
P_{MP}	Polynomial order in an MP model
P_{V}	Polynomial order in a Volterra model
P_{W}	Polynomial order in a Wiener model
$\mathbf{q}^{\mathrm{re}}$	Real-valued spline control point set of the I branch
$\mathbf{q}^{\mathrm{im}}$	Real-valued spline control point set of the Q branch
$\mathbf{q}_n$	Complex control points of the spline interpolation scheme
$\mathbf{q}_{m,n}$	Control points of the spline interpolation scheme used in SPMP
$\mathbf{Q}$	Conjugated covariance matrix
Q^{SP}	Number of control points in a spline-interpolated scheme
$r(t)$	Received useful signal in an IBFD transceiver
$r[n]$	Received useful signal in an IBFD transceiver after the ADC
$\mathbf{R}$	Input data covariance matrix
$\hat{\mathbf{R}}$	Classical sample estimation input data covariance matrix
$\mathbf{R}_k$	kth $M \times M$ sub-block matrix in the ICM estimation methods
R	No. of antennas and PAs in an active antenna array transmitter
R_{X}	Autocorrelation function with respect to the variable x
R_{XZ}	Cross-correlation function of the variable x with respect to z
$s[n]$	First intermediate signal in a cascaded system
$\mathbf{s}_n$	Signal regression of $s[n]$
$SINR$	Signal-to-noise plus interference ratio
t	Time variable
t_k	kth knot within a knot set
$\mathbf{T}$	Knot set
u_n	Abscissa value in the spline interpolation scheme at time instant n
$\mathbf{u}_n$	Abscissa vector in the spline interpolation scheme at time instant n
$\dot{\mathbf{u}}_n$	Derivative of the abscissa vector in the spline interpolation scheme

$w[k]$	kth linear FIR filter coefficients of $\mathbf{w}_n$
$\mathbf{w}_n$	Additional linear FIR filter coefficients used in SPWH
$x[n]$	Input signal in an SISO transmitter
$\tilde{x}[n]$	Signal after DPD main path processing
$\mathbf{x}_k$	kth input sample vector
$\mathbf{x}_n$	Signal regression of $x[n]$
$\mathbf{x}_n^{\mathrm{BF}}$	nth row of the input data BF matrix $\mathbf{\Omega}$
$x_{\mathrm{RF}}(t)$	Upconverted signal after the DAC
$x_{\mathrm{SI}}(t)$	Received SI signal in an IBFD transceiver
$x_{\mathrm{SI}}[n]$	Received SI signal in an IBFD transceiver after the ADC
$\hat{x}_{\mathrm{SI}}[n]$	Estimated SI signal in an IBFD transceiver
$x_{\mathrm{PA}}(t)$	Output TX signal in an IBFD transceiver
$\dot{\mathbf{X}}_n$	Diagonal matrix containing the signal regression of $x[n]$
$y[n]$	Complete output signal of a digital signal model
$\dot{y}[n]$	Derivative of the spline-interpolated output signal
$y_{\mathrm{H}}[n]$	Output model signal in a Hammerstein system
$y_{\mathrm{L}}[n]$	Output model signal in an LUT system
$y_{\mathrm{MP}}[n]$	Output model signal in an MP system
$y_{\mathrm{NL}}[n]$	Output model signal in a nested LUT system
$y_{\mathrm{V}}[n]$	Output model signal in a Volterra system
$y_{\mathrm{W}}[n]$	Output model signal in a Wiener system
$\mathbf{y}_k$	kth output sample vector
$\dot{\mathbf{Y}}_n$	Diagonal matrix containing the signal regression of $\dot{y}[n]$
$z[n]$	Observed received signal after the ADC
$z(t)$	Overall received signal in an IBFD transceiver
$z_{\mathrm{n}}[n]$	Noise signal after the ADC
$z_{\mathrm{n}}(t)$	Noise signal

$z_{\mathrm{RF}}(t)$	Observed PA output signal before the ADC
$\boldsymbol{\alpha}$	Complete set of control point coefficients in the MP-LUT model
α_k	kth element of the control point set $\boldsymbol{\alpha}$
$\gamma(\cdot)$	Output of the spline interpolation scheme
$\mathbf{\Gamma}_n$	Computing matrix in the SPWH model, at time instant n
Δ_x	Region width in the interpolated scheme, with input signal $x[n]$
∇_x	Gradient operator with respect to the variable x
ζ	Permutation of an arbitrary set of integers
μ_{x}	Learning rate associated to the learning equation of x
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ, and variance σ^2
ξ_k	kth coefficient of the Bussgang's method
$\mathbf{\Xi}$	Bussgang's coefficient matrix
σ_x	Signal variance of the signal x
$\mathbf{\Sigma}_n$	Computing matrix in the SPWH model, at time instant n
τ	An arbitrary sample delay
τ^{ICM}	Reduced sample-span of the recursive ICM estimation method
τ^{SP}	Reduced time-span considered in SPH, SPWH, and SPHW models
$\boldsymbol{\upsilon}[n]$	Instantaneous input data vector, first row of $\mathbf{\Omega}$
$\upsilon_k[n]$	kth element of the instantaneous input data vector
ϕ	An arbitrary frequency shift
$\mathbf{\Phi}$	Control point weight in the spline interpolation scheme
$\mathbf{\Psi}$	Control point weight in the spline interpolation scheme
$\phi\omega$	Normalized angular frequency, in Hz
$\mathbf{\Omega}$	Complete input basis function matrix
$\mathbf{\Omega}_k$	K^{it}-sized input basis function matrix at iteration k

1 INTRODUCTION

1.1 Thesis Motivation and Scope

DURING recent decades, there has undoubtedly been a massive increase in the usage of wireless communications. A vast majority of devices surrounding us have some sort of wireless data transmission, from our personal phone to our refrigerator, which automatically orders food by itself when the reserves are low. The exponential demand for new applications, higher peak data rates, reliability, and a massive number of devices connected together call for new technological solutions which utilize more efficiently the available radio spectrum and energy resources. On the one hand, increasing the spectral efficiency is a key aspect which has been given special focus in recent years [10], [11], [30], [59], and this is particularly so due to the physical limitation wireless communications suffer from. This means that the available radio spectrum has a certain range of frequencies which can be utilized for wireless transmission, especially for wide-area networks, and a time will come when all these usable frequencies will have been filled. It is thus crucial to really exploit each frequency band to provide the services future society will require. On the other hand, however, increasing the energy efficiency of the transmitters drastically reduces the circuitry operational costs and production size. Firstly, decreasing the associated costs benefits the telecommunication operators and component manufacturers, who are able to offer their products at a more reduced prize. This further eases the deployment of machine-to-machine (M2M) communication, particularly sought within modern fifth generation (5G) networks [61], [121], [216], facilitating having a massive number of machine-type communication (MTC) nodes connected together. Additionally, this also helps to connect people at an inter-personal level. Secondly, reducing the size of the circuitry involved reduces the carbon footprint, helping the environmental sustainability of our planet. The overall objective of this thesis is then to provide several solutions which can potentially bring more efficient

spectrum utilization and more efficient transmitters one step closer to commercially feasible implementations.

In terms of energy efficiency of any transmitter, the clear bottleneck is typically found in the transmit power amplifier (PA). In order to ideally achieve maximum energy efficiency from such a component, it needs to be operated in a highly saturated region, or equivalently, by providing maximum output power. This, however, is not feasible in reality, as any PA presents a nonlinear input-output power response that heavily distorts its output signal. This is particularly problematic with modern radio communication systems, which typically build on multicarrier modulations, for instance, orthogonal frequency division multiplexing (OFDM) [131], [237], [269]. Such waveforms are known to contain high values of peak-to-average power ratio (PAPR), which further complicates the use of highly saturated PAs, as the high powered peaks of the signal fall into the saturated region, producing nonlinear distortion at the PA output [247]. Any PA typically creates three main types of nonlinear distortion: i) in-band distortion, which affects the quality of the signal itself and degrades the normalized mean square error (NMSE) and error vector magnitude (EVM); ii) out-of-band (OoB) distortion, which is seen as a spectral regrowth overlapping the adjacent channels in the frequency spectrum, and degrades the adjacent channel leakage ratio (ACLR), or the ACLR measured with the total radiated power (TRP) approach [1]; and iii) spurious emissions, which may fall over other bands of interest beyond the adjacent channel. The maximum allowed levels of these unwanted emissions are enforced by several associations at national and international levels, such as the 3rd generation partnership project (3GPP) or the international telecommunications union (ITU), among others. Any commercial wireless transceiver needs to comply with these regulations, hence exploring ways of minimizing unwanted emissions is of great importance.

One plausible solution to reduce these types of unwanted distortion is to back off the PA, or in other words, to operate it in a less saturated region, thus reducing its output power. Doing so, however, will noticeably decrease the energy efficiency of the PA, thus this is not considered as an attractive approach at all. A more appealing solution would be to use some kind of PA linearization method, which ensures that a high power efficiency is maintained in the PA while keeping the unwanted in-band and OoB distortion below specified thresholds. Digital predistortion (DPD) is a well-known and a widely applied PA linearization technique,

which has been utilized since the early 90s, and has been proven to provide reliable and flexible performance [91], [193], [237]. DPD aims at applying a digital nonlinear pre-compensation to the transmit PA waveform that compensates the nonlinear response of the PA. Many DPD algorithms have been proposed by the scientific community, however, the research into low-complexity DPD is more scarce.

Low complexity DPD algorithms are particularly sought after in modern 5G NR networks [10], [48], [90], and this is particularly motivated by the following facts. First, the signal bandwidths in 5G are substantially higher compared to long-term evolution-advanced (LTE-A) [60], [120], [123]. 5G NR considers bandwidths up to 100 MHz in frequency range 1 (FR-1) (below 6 GHz bands), and up to 400 MHz in frequency range 2 (FR-2) (24-40 GHz bands) [1], [44], [196], which radically increases also the associated DPD processing rates. Second, small base station (BS) configurations, such as *medium range* or *local area* BSs, constitute a new trend in modern networks, with associated reduced transmit powers. This translates also to a reduced power budget for the required front-end signal processing techniques. Third, in modern FR-2 systems, the ACLR and TRP ACLR OoB emission requirements are largely relaxed, facilitating the deployment of simpler DPD solutions. Fourth and final, solutions providing continuous parameter tracking are frequently pursued, [30], [85], [138], [207], especially in FR-2 systems and other millimeter wave (mmW) beam-steered multi-active antenna arrays, thus reducing the complexity of the parameter learning stage also becomes important.

With respect to enhancing spectral efficiency and/or data rates, several techniques have been proposed. In the recently launched 3GPP's release 15 technical specification, a special emphasis is given to high frequency communication, where the available spectrum is still not scarce [85], [204], [212]. Another plausible approach is the adoption of so-called multiple-input multiple-output (MIMO) transmission, which benefits from multiple transmit/receive antennas capable of providing spatial multiplexing through more than one stream of useful information [6], [111], [158], [175], [190]. An alternative solution which has gathered a lot of attention during the past decade is the potential use of in-band full-duplex (IBFD) technology, which constitutes, in part, the main topic of this thesis. An IBFD transceiver is theoretically capable of transmitting and receiving useful information in the same frequency and at the same time, without requiring any additional bandwidth [30], [51], [128]. This translates to a twofold increase in terms of spectral efficiency and data rates, since the

full-duplex transceiver can operate all the time, in contrast to a half-duplex system where the time or frequency resource needs to be shared between transmission and reception. This is the case in time-division duplex (TDD) or frequency-division duplex (FDD) systems.

IBFD is a clear candidate to be implemented in future wireless networks, but also has associated challenges. The inherent problem of IBFD is the so-called self-interference (SI), which occurs when the system's own transmit signal is leaked to the receiver chain of the IBFD system itself. This interference signal can be many orders of magnitude stronger than the received useful signal, thus it needs to be removed by some means in order to make IBFD commercially feasible [30], [77], [217]. In theory, the SI can be cancelled by just removing a replica of the transmit signal, properly scaled, from the overall received signal. However, in reality, this process becomes much more complicated, stemming from the distortion that is injected by all the transmitter (TX) and receiver (RX) circuitry of the transceiver, and also stemming from the random multipath response of the coupling channel between the transmitter and receiver. A digital SI canceller must take these effects into account, hence these algorithms need to first *model* all the responses affecting the SI, to then properly *cancel* its effect from the overall received signal, such that only the pure useful signal contribution remains. Since any IBFD transceiver needs to necessarily remove the SI contribution, it is crucial to explore ways of doing it by minimizing the involved computational complexity.

In summary, this thesis focuses on the development of low-complexity DPD and digital SI cancellation techniques to improve radio communication energy-efficiency and spectral-efficiency with computationally feasible solutions. On the one hand, DPD enables operating the PA within a highly nonlinear region, thus increasing its power efficiency, while keeping the unwanted in-band and OoB emissions within specified levels. On the other hand, IBFD aims at providing a twofold increase in terms of spectral efficiency and data rates by transmitting information in the same frequency band and at the same time, avoiding the need to share the time/frequency resources between the transmitter and the receiver. These questions are currently under intensive study by the scientific community, and also constitute the main topics of this present thesis.

1.2 Thesis Objectives

This thesis work focuses on both DPD and IBFD applications, but there is a common clear objective between both, which is the design of more energy-efficient transmitters, with associated lower production costs and size. These characteristics reduce operational and production costs, and also help in reducing the carbon footprint and hence our environment. An enhanced energy efficiency is achieved, in part, by utilizing more efficient hardware components in the TX/RX chain of the communication devices, and more specially by using more nonlinear PAs which can provide as high transmit powers as possible. Energy efficiency can also be achieved by optimizing the required algorithms which need to be executed within the device, in order that it drains as little processing power as possible.

First, in the case of DPD, the goal is to develop novel structures and techniques to linearize such challenging nonlinear scenarios, while minimizing the complexity. The DPD operation can be divided in two main stages: firstly, the predistorter main path, which needs to be continuously executed to predistort the input signal, and thus a particularly low complexity is desired; and secondly, the DPD learning path, which is required to perform the parameter estimation, where simple techniques are sought in order to provide fast DPD tracking and real-time DPD running. Finally, to fulfill this goal, the performance of the proposed reduced-complexity techniques is analyzed through real-life radio-frequency (RF) measurements, and compared to currently applicable regulatory requirements to show that they are indeed fulfilled, regardless of the enhanced energy efficiency achieved in the transmit chain.

Second, in the case of IBFD, any device must inevitable perform the SI cancellation in order to successfully operate, hence it is key to explore different ways of minimizing the involved complexities, while still providing sufficient levels of SI cancellation. The objective of this thesis is then to study different architectures and approaches capable of minimizing the involved complexity in the digital cancellation stage, which any IBFD device must perform. The proposed solutions are evaluated through real-life RF measurements, which involve the use of realistic IBFD prototypes. The results suggest that the digital SI cancellation stage can be optimized in order to reduce the involved complexity, while maintaining sufficient levels of SI cancellation – even when operating under a energy-efficient and heavily nonlinear transmitter power amplifier.

1.3 Thesis Contributions and Structure

The main contributions of this thesis can be summarized as follows:

- Four different spline-interpolated cascaded modeling architectures are proposed, alongside with their corresponding decoupled gradient-based adaptive learning solutions, in [P1, P3, P5, P7], and [21], [146], [155]. Three of these architectures are used in the context of DPD and PA linearization, and the remaining architecture is utilized in the context of digital SI cancellation.
- A linear-in-parameters spline lookup-table memory polynomial (MP-LUT) modeling architecture is derived, alongside with its corresponding gradient adaptive learning solution [P1, P2, P4, P6, P8]. This model incorporates a spline interpolation scheme in each memory polynomial (MP)-like parallel branch, thus the interpolation order can be modified as desired to result in no interpolation, linear interpolation, or higher-order interpolation.
- A new formulation of the DPD main path injection-based processing scheme for models based on look-up tables (LUTs) is presented in [P1-P8] and [21], [146], [155], [200]. This scheme removes gain ambiguities in the cascaded modeling systems, and reduces the dynamic range of the LUT in all cases [P2, P3], as only the deviation from unit gain is coded in the LUTs.
- Several reduced-complexity closed-loop sign algorithms are applied to classical Gauss-Newton (GN), self-orthogonalized (SO), and block-least mean squares (BLMS) learning solutions [P2, P6]. Specifically, the sign algorithm, sign-regressor algorithm (SRA), and sign-sign algorithm are utilized, significantly reducing the overall learning computational complexity.
- Several low-complexity inverse covariance matrix (ICM) estimation methods for the SO learning solution are derived [P4, P8]. These methods focus on the i) estimation of ICM from the autocorrelation function of the input data, ii) approximation of the ICM by using Bussgang's coefficients, iii) reduction of the dimensionality of the ICM, iv) estimation of the ICM by applying a frequency shift, and v) recursive estimation of the ICM.
- Detailed complexity analyses and comprehensive real-life RF measurement results of all proposed techniques in [P1-P8] and [21], [68], [155], [200] are

presented throughout this thesis, which offer proof and verification of the achieved performance and complexity reductions. This also allows for extracting the performance-complexity trade-offs of each proposed method.

This thesis work presents a summary of the proposed techniques in [P1-P8], thus only essential information and results are included herein. For a more complete explanation of the system models, complexity analyses, and results, the reader is kindly referred to the aforementioned publications. For clarity and cohesion, a slightly different notation from the original publications is utilized within this thesis.

The thesis is organized as follows. Chapter 2 overviews modern challenges in wireless communications, and presents the basics of DPD and IBFD. In addition, a state-of-the art summary of both technologies is presented. Chapter 3 first develops the complex-valued, injection-based spline interpolation scheme, and follows with the development of all the spline-interpolated cascaded models, alongside with their corresponding gradient-descent adaptive learning equations. Chapter 4 overviews the spline-based MP-LUT architecture, and, in the context of linear-in-parameters models, introduces also the classical closed-loop GN, SO, and BLMS learning equations. In addition, the reduced-complexity sign-based algorithms and ICM estimation methods are reviewed. All the proposed techniques are then evaluated in Chapter 5, which provides comprehensive real-life RF measurements to test and verify the proposed solutions. Finally, the main findings of this thesis and possible future directions are presented in Chapter 6.

1.4 Author's Contributions to the Publications

The research field and topic ideas carried out throughout this thesis were originally proposed by Prof. Mikko Valkama. He has collaborated in all the publications made by the Author of this thesis (later: the Author) by sharing his expertise, not only in the associated technical parts, but also in the writing process. In addition, Dr. Lauri Anttila, who has co-supervised this thesis, has also widely cooperated through the research process, providing very useful technical ideas and extensive writing feedback.

The publication contained in [P5] served as a starting point for this thesis work, in which Dr. Dani Korpi taught the Author the basics of IBFD technology, in terms of mathematical derivations and RF experimentation. Thus, this publication was

in part written by Dr. Dani Korpi, and in part written by the Author, while Prof. Mikko Valkama and Dr. Lauri Anttila assisted with the writing process.

In [P1, P3, P7], regarding reduced-complexity solutions for DPD and IBFD, the Author of the thesis performed all the mathematical derivations, system simulations, and experimental RF measurements. In [P1], Dr. Lauri Anttila provided useful ideas regarding the injection-based method applied in the spline interpolation scheme, and also helped to develop more efficient spline interpolation processing. In addition, M.Sc. Alberto Brihuega provided the initial basis regarding the configuration of the FR-2 setup, utilized to carry out one set of experimental measurements at mmW frequencies, as it was the first time the Author had used it. Other authors helped with the final appearance of the paper. In [P3], M.Sc. Matias Turunen greatly helped at the time of configuring the RF measurement setups which were utilized to carry out the RF experimentation, although all the measurement results were carried out by the Author. Finally, in [P7], M.Sc. Alberto Brihuega provided the idea of using a metallic reflector to obtain the SI signal in the RX antenna. It is finally noted that Prof. Mikko Valkama and Dr. Lauri Anttila helped with the writing process in all the aforementioned publications.

The works contained in [P2, P4, P6, P8] were carried out in collaboration with Huawei technologies Sweden AB. All the mathematical derivations and RF verification were carried out by the Author, while other collaborators provided useful ideas which helped improving the quality of the publications. Specifically, in [P2] and [P6], Dr. Guo Yan introduced the idea of using the SRA applied to block-based learning equations to reduce their learning path complexities. Finally, the work in [P8] served as a starting point for [P4], where Dr. Lauri Anttila proposed the initial research direction in Bussgang's and autocorrelation inverse covariance estimation methods. In addition, Dr. Neng Wang proposed the used of the Kronecker notation in the Bussgang inverse covariance matrix estimation method to reduce the inversion complexity.

In addition to these publications, the Author has co-authored a book chapter [146], authored another journal paper [200], and co-authored the works presented in [21], [68], [98], [155], [244], [245]. These research works are not attached to this thesis, but constitute also part of the research activity carried out by the Author during his research time.

1.5 Nomenclature

Throughout this thesis, matrices are represented by capital boldface letters, e.g., $\mathbf{A} \in \mathbb{C}^{M\times N}$, and vectors are column sets, presented with lowercase boldface letters, e.g., $\mathbf{v} \in \mathbb{C}^{M\times 1} = [v_0\ v_1\ \cdots\ v_{M-1}]^T$, where the subindex indicates the entry index. By default, matrices and vectors are assumed to be complex-valued, unless otherwise mentioned, with ordinary transpose, Hermitian transpose, and complex conjugation denoted by $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. A signal regression is represented as a column vector containing its M past samples, e.g., $\mathbf{x}_n \in \mathbb{C}^{M\times 1} = [x[n]\ x[n-1]\ \cdots\ x[n-M+1]^T$. Additionally, the expected value, absolute value, floor, ceil, factorial, convolution, Hadamard product, and Kronecker product operators are written as $\mathrm{E}\{\cdot\}$, $|\cdot|$, $\lfloor\cdot\rfloor$, $\lceil\cdot\rceil$, $!$, $\star$, $\circ$, and $\otimes$, respectively.

Furthermore, recursive algorithms have an additional discrete-time variable, indicated as a subindex. In the left part of the equality, this subindex is incremental, indicating the value corresponds to the next algorithm iteration. Considering an additional learning rate μ_{x}, an example of a complete recursive law reads

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_{\mathrm{x}} \cdots . \tag{1.1}$$

Finally, regarding the steepest descent gradient algorithm, the gradient operator is expressed as ∇_x, indicating that the partial derivative is taken with respect to the complex conjugate of the parameter x, i.e.,

$$\nabla_x J(\cdot) = \frac{\partial J(\cdot)}{\partial x^*}. \tag{1.2}$$

2 BACKGROUND AND STATE OF THE ART

THIS chapter first provides an overview of emerging trends in modern wireless communication systems, while also presenting their associated challenges and diverse possibilities of dealing with them. After this, essential background information regarding both digital predistortion (DPD) and in-band full duplex (IBFD) applications is provided, together with a detailed state-of-the-art summary containing the most relevant publications in these fields, as per the Author's knowledge.

2.1 Trends and Challenges in Modern Wireless Communications

Wireless communications are evolving following giant steps, facilitating services and applications that have an impact in every aspect of our lives. These services and applications require more stringent requirements as technology progresses, such that old LTE-A networks are already struggling to satisfy these demands, and are basically reaching their performance boundaries. In the second quarter of 2019, 5G new radio (NR) Release 15 was officially launched by the 3GPP [1], [61], which promised higher peak data rates, increased uplink (UL) and downlink (DL) bandwidths (up to 400 MHz in FR-2, in contrast to 20 MHz in LTE), ultra reliable low-latency communications (URLLC) and MTC, and a massive network capacity expansion, among many others. 5G NR also considers a slightly modified network architecture, oriented towards small-cell configuration which will support 100 times more users than in LTE-A, in which this number was limited to 200-400 users per cell. All of these new network aspects have a direct impact on the utilized signal processing practices, both in the BS and in the user equipment (UE). Reduced-complexity techniques are a new trend in 5G NR [61], [69], [220], which are pursued to increase the power-efficiency of the BS/UE. These techniques are mainly motivated by the following

facts or tendencies: i) As mentioned before, the signal bandwidths in 5G NR are greatly increased with respect to LTE-A [34], [95]. 5G NR considers bandwidths up to 100 MHz in FR-1 (below 6 GHz bands), and up to 400 MHz in FR-2 (24-40 GHz bands) [1], which radically increase also the associated processing rates; ii) in FR-1 frequencies, a *medium range* and *local area* BS will radiate less transmit powers compared to an old macro BS, thus the available power budget for the required signal processing techniques will also be reduced; iii) in FR-2 systems, the ACLR and TRP ACLR OoB emission requirements are largely relaxed, being in the order of just +26-28 dBc, facilitating the deployment of simpler DPD solutions [153]; iv) finally, as demonstrated in [30], [138] in the context of IBFD, and in [85], [129], [207] in the context of DPD, continuous parameter tracking may be needed, especially in FR-2 systems and other mmW active antenna arrays, where the electrical beam is steered, thus reducing the complexity of the parameter learning stage becomes also important.

2.1.1 PA Nonlinear Distortion Quantification

When a PA is operated in its saturated region, its output becomes distorted. Nevertheless, it is desirable to operate PAs in such regions, as the power efficiency increases with output power. The high PAPR of modern multicarrier waveforms complicates even more the utilization of highly nonlinear power amplifiers [55], [210], [269], as the instantaneous high peaks of the input signals fall into a more saturated region, producing larger amounts of nonlinear distortion at its output [37], [55]. The said output nonlinear distortion can be characterized in three main types. The first type is in-band distortion, which affects the quality of the signal itself, and can be reflected in the NMSE and EVM performance metrics. The second type is OoB distortion, which is seen as a spectral regrowth overlapping the adjacent channels, hence causing potential disruption and interference to adjacent channel users. The OoB effect is reflected in the ACLR or TRP ACLR performance metrics. The third kind of distortion is seen as instantaneous spurious responses that may fall over other bands of interest which are beyond the adjacent channel.

This thesis focuses on compensating for the in-band and OoB nonlinear distortion, while it is assumed that the spurious emissions can be compensated by means of a filtering stage that attenuates the regions beyond the adjacent channels. When refer-

ring to the OoB distortion, the ACLR metric is adopted to measure the performance in classical wired single-input single-output (SISO) systems, while the TRP ACLR metric is adopted to quantify the performance in FR-2 systems where an over-the-air (OTA) active antenna array is deployed [1], [85], [191]. These metrics are explained in the following lines.

In-band Nonlinear Distortion Quantification

The in-band distortion can be quantified through the well known NMSE and EVM performance metrics [97], [156], [237]. These metrics can be defined as

$$\text{NMSE (dB)} = 10\log_{10}\left(\frac{P_{\text{error}}}{P_{\text{ref.}}}\right), \tag{2.1}$$

$$\text{EVM (\%)} = \sqrt{\frac{P_{\text{error, eq.}}}{P_{\text{ref.}}}} \times 100, \tag{2.2}$$

where P_{error} denotes the power of the error signal, which is calculated as the difference between the measured and reference signal, and $P_{\text{ref.}}$ denotes the power of the reference signal. In the case of EVM, $P_{\text{error,eq.}}$ corresponds to the power of the error signal, but is now calculated between the ideal subcarrier symbols and the corresponding measured subcarrier samples at the PA output, after zero forcing equalization, which removes the effects of the linear distortion [60], [229], [271], and $P_{\text{ref.}}$ denotes the power of the reference symbols. In this thesis work and in publications [P1-P8], both the NMSE and EVM are used to evaluate the inband distortion. The NMSE and EVM are statistically equivalent [156], [237].

Out-of-band Nonlinear Distortion Quantification

The OoB distortion can be quantified, in turn, by using the well-known ACLR, the metric which measures the ratio between the transmitted power in the desired channel and the spilled power in an adjacent channel [113], [271]. The ACLR can be presented for each left and right adjacent channels, or alternatively left as the lowest value (i.e., most restrictive value) obtained out of those. Formally, the ACLR can be

expressed as

$$\text{ACLR (dB)} = 10\log_{10}\Big(\frac{P_{\text{desired ch.}}}{P_{\text{adj. ch.}}}\Big), \tag{2.3}$$

where $P_{\text{desired ch.}}$ is the power within the desired channel, and $P_{\text{adj. ch.}}$ is the power within the adjacent left or right channels. The TRP ACLR metric is used when measuring the spill-over power of OTA systems. The TRP ACLR metric is equivalent to the classical ACLR, but it is measured by integrating the powers over the whole beamspace of the antenna array, while keeping the *beamforming angle* fixed [1], [42], [85]. For simplicity, the ACLR metric, measured with the TRP method, is referred to as simply TRP ACLR in this thesis.

2.1.2 SI Quantification in IBFD Applications

The inherent problem of IBFD technology is the self-interference (SI) leakage. In other words, a device which is transmitting at the same time and frequency will automatically induce a signal in its own RX chain which is many orders of magnitude more powerful than the actual desired received signal. In some cases, the induced SI has been reported to be more than +100 dB stronger than the desired signal [150], [217]. In order to quantify the required SI in a generic 5G NR IBFD transceiver, let's consider an example modern 5G NR UE, with a maximum transmit power of +20 dBm and a sensitivity of -90 dBm [202], [222]. This means that the TX-RX isolation should be at least +110 dB in order to suppress the SI to the receiver noise floor level. Another example is a *local area* BS, with a maximum transmit power of +24 dBm and a minimum sensitivity of -93.7 dBm, in the most restrictive scenario. With this configuration, the required SI suppression is in the order of +118 dB in order to suppress the SI to the receiver noise floor level. The overall effect of the SI signal is reflected as a decrease in the signal-to-interference plus noise ratio (SINR) achieved by the IBFD system, and this translates to a reduction in the achievable channel capacity. This concept is given by the Shannon Hartley theorem [215], which relates the channel capacity, C_{IBFD}, the signal bandwidth, BW, and the SINR of the system, as

$$C_{\text{IBFD}} = 2BW\log_2(1+SINR). \tag{2.4}$$

Note that the "2" multiplier in this expression appears only in the IBFD case, where the capacity is in theory doubled due to the simultaneous transmission and reception (STAR) of the IBFD transceiver. It is also noted that this expression assumes a Gaussian distribution for the interference plus noise signals, thus it constitutes an approximation if other non-Gaussian waveforms are utilized. The concept of SI is further explained in Section 2.3.1.

Since any IBFD transmitter must suppress the SI for proper operation, it is crucial to explore ways that minimize the involved computational complexity, while still achieving highly accurate SI modeling performance, which fulfills in all cases the IBFD transceiver requirements. In theory, the SI can be cancelled pretty easily. Since the transmitter obviously knows its own transmit signal, the overall process could be done by just subtracting a replica of it, properly scaled, from the overall received signal. Of course, this would only happen if assuming that the TX and RX chains present a completely linear response. However, in practice, all the circuitry components of the TX and RX chains are inducing nonlinear distortion to the transmit signal, and also the multipath channel between the TX and RX makes a huge contribution to it. The overall received SI contribution is thus heavily distorted by several unknown responses, and this is what makes the problem of SI cancellation so challenging. Typically, the elements contributing the most to SI distortion are the TX PA and the SI channel. The TX PA is usually operated in its saturated region to ensure a sufficiently high power efficiency in the transmitter, but this induces heavy nonlinear distortion to its output signal, as already discussed earlier in this section. The SI channel typically presents random multipath effects, which can be compensated by means of elementary linear time-invariant (LTI) filters. There are also many other impairments contributing to the SI distortion, such as the in-phase/quadrature (I/Q) imbalance induced by the TX and RX mixers, and the nonlinear distortion of the low-noise amplifier (LNA), among others.

In order to provide a reliable and accurate replica of the SI, digital cancellation algorithms must take these unwanted effects into account. Hence, such techniques should *model* the response of the distorting elements in the TX/RX chains of the transceiver, and *cancel* them from the overall received signal in a posterior digital cancellation stage, such that only the pure useful received signal remains.

2.2 Digital Predistortion

The transmit RF PA is located before the antenna in any transmit communication system, and it is a crucial element to ensure enough power is transmitted. The PA is either linear or efficient, its efficiency increasing with its output power [58], [269]. Unfortunately, as the PA output power increases, so does its nonlinear response, thus high levels of distortion are induced to the passing waveforms. The de-facto solution to cope with this issue is DPD. DPD applies a nonlinear pre-transformation to the input digital waveform capable of compensating for the in-band and OoB distortions created by the nonlinear PA. Generally, any DPD system is divided into two stages: the DPD main path, which applies the predistortion to the input digital signal, and the DPD learning path, which estimates and updates the DPD coefficients.

2.2.1 Basics of DPD

In the following sections, the essential and basic background of DPD is reviewed, including fundamental nonlinear models and system architectures. This background will facilitate the development of the posterior proposed DPD models and solutions, with the aim of reducing the involved computational complexity.

Nonlinear Models for Digital Predistortion

In general, PA nonlinear models can be divided into two types: physical models and behavioral models [58], [177], [187], according to the available physical knowledge of the PA internal circuitry. On the one hand, physical models take into account the internal PA equivalent circuit, with associated current and voltages to create a very accurate nonlinear equivalent circuit description. This method provides accurate results, however, it is often very time consuming and computationally heavy. On the other hand, behavioral models consider the PA as a black-box model, just relying on its input-output relation to extract the DPD coefficients. These models are generally less accurate than physical models, and their precision will also depend on the chosen parameter estimation procedure [91], [177]. In this thesis, the latter approach is adopted, since it provides simpler solutions which are more suitable for reduced-complexity and real-time or fast-tracking DPD solutions.

Polynomial-based approaches are the most widely used behavioral models for PA model identification [91], [133], [237]. These models are known to be linear-in-parameters [91], [134], [270], thus they constitute simple solutions for DPD, and also have good approximation capabilities with the extra possibility of adding or removing memory terms as desired [50], [136], [248]. The most general polynomial-based model is the Volterra approach, also known as the Volterra series [29], [104], [267], [268]. Considering a polynomial order P_{V}, a memory depth M_{V}, and naming the Volterra kernels as $f_{p,m_1,m_2,\cdots,m_p}$, the discrete time input-output relation of this model reads [177], [237]

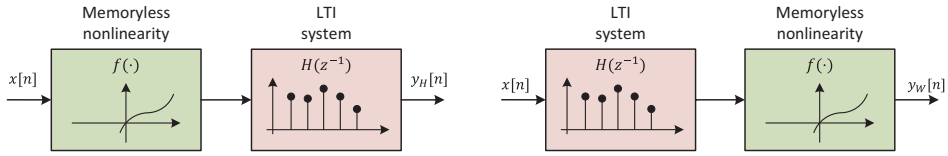
$$y_{\mathrm{V}}[n] = \sum_{\substack{p=1 \\ p\,\text{odd}}}^{P_{\mathrm{V}}} \sum_{m_1=0}^{M_{\mathrm{V}}} \sum_{m_2=m_1}^{M_{\mathrm{V}}} \cdots \sum_{\frac{m_{p+1}}{2}=\frac{m_{p-1}}{2}}^{M_{\mathrm{V}}} \cdots \sum_{\frac{m_{p+3}}{2}=0}^{M_{\mathrm{V}}} \cdots \sum_{m_p=m_{p-1}}^{M_{\mathrm{V}}} f_{p,m_1,m_2,\cdots,m_p} \times \prod_{i=1}^{\frac{p+1}{2}} x[n-m_i] \prod_{k=\frac{p+3}{2}}^{p} x^*[n-m_k], \tag{2.5}$$

where only odd-order nonlinear products are considered [70], [104], [177]. Unfortunately, the number of basis functions (BFs) of this model is massive, and grows exponentially with the nonlinearity order and memory depth, thus it is unfeasible to use without any complexity-reduction technique, as the ones presented in [38], [266], [272]. Also, different subsets of the Volterra series with a reduced number of BFs can be used, such as the generalized memory polynomial (GMP) model [174], [188], [193], or the classical MP model, whose input-output relation simply reads [74], [135], [237]

$$y_{\mathrm{MP}}[n] = \sum_{m=0}^{M_{\mathrm{MP}}-1} \sum_{\substack{p=0 \\ p\,\text{odd}}}^{P_{\mathrm{MP}}} \alpha_{m,p}\, x[n-m]|x[n-m]|^{(p-1)}, \tag{2.6}$$

where P_{MP} is the polynomial order [82], [157], M_{MP} is the memory depth, and $\alpha_{m,p}$ are the MP coefficients.

There are also several cascaded structures that can be used to model the nonlinear behavior of the PA, the two most common examples being the Hammerstein and Wiener systems, conceptually presented in Fig. 2.1. The Hammerstein system incorporates an instantaneous nonlinear block followed by an LTI elementary block,



(a) A Hammerstein architecture.

(b) A Wiener architecture.

Figure 2.1 A conceptual illustration showing the block diagram of the (a) Hammerstein system model, and (b) Wiener system model.

which ideally models a system whose structure is alike. The elementary blocks in the Wiener system are inverted, as shown in Fig. 2.1b.

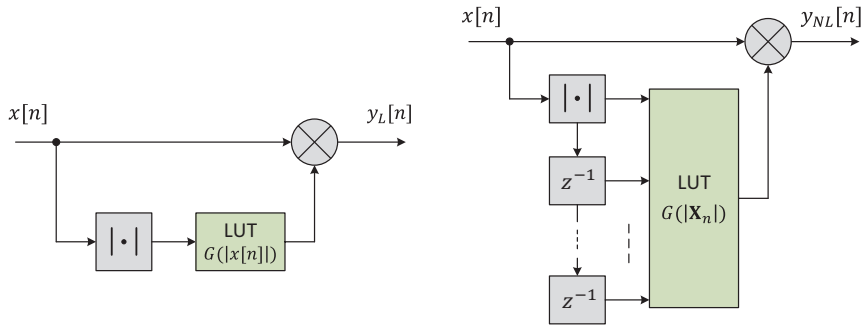
Assuming that the memoryless nonlinearity is implemented with a polynomial expression, and the LTI block with an finite impulse response (FIR) filter, the input-output relation of the Hammerstein and Wiener models read [92], [177], [224], [225], respectively

$$y_{\mathrm{H}}[n] = \sum_{m=0}^{M_{\mathrm{H}}} w_m \sum_{p=0}^{P_{\mathrm{H}}} \alpha_p x[n-m]|x[n-m]|^p, \tag{2.7}$$

$$y_{\mathrm{W}}[n] = \sum_{p=0}^{P_{\mathrm{W}}} \alpha_p \Big(\sum_{m=0}^{M_{\mathrm{W}}} w_m x[n-m] \Big) \Big| \sum_{m=0}^{\mathrm{W}} w_m x[n-m] \Big|^p, \tag{2.8}$$

where w_m are the LTI filter coefficients, and α_p are the polynomial nonlinear coefficients. It is noted that many other cascaded structures can be used for DPD, such as the Hammerstein-Wiener or Wiener-Hammerstein, which are studied later in this thesis. Cascaded approaches typically have a lesser number of adaptation coefficients, thus their complexity is generally decreased with respect to linear-in-parameters polynomial-based models with memory [94], [166], [225]. However, they require *physical knowledge* of the system, which is not always known.

Another alternative approach to model the nonlinear PA is to use LUT-based models [130], [163], [173], [207], [264]. Out of these, the two most common approaches are the simple LUT and the nested LUT models, presented conceptually and respectively in Fig. 2.2. The simple LUT method – Fig. 2.2a – forms the basic behavioral model for memoryless PA nonlinearities, and simply applies a complex-



(a) An LUT model. **(b)** A nested LUT model

Figure 2.2 A conceptual illustration showing the block diagram of a (a) LUT model, and (b) nested LUT model.

value gain to the input waveform as

$$y_{\mathrm{L}}[n] = G(|x[n]|)x[n], \tag{2.9}$$

where $G(|x[n]|)$ is the complex gain, which depends on the magnitude of the input signal. Such complex-valued gain is generally stored in two real-valued LUTs. On the other hand, the nested LUT scheme – Fig. 2.2b – was proposed to enhance the modeling capabilities of the former LUT model by adding memory effects [63], [114], [243]. To this end, the complex LUT gain is now a function of the instantaneous input sample plus an additional $M_{\mathrm{L}} - 1$ preceding samples. This way, M_{L} is the total amount of memory considered in the model, and its input-output relation reads

$$y_{\mathrm{NL}}[n] = G(|\mathbf{X}_n|)x[n], \tag{2.10}$$

where now the complex-valued gain $G(\cdot)$ is a function of the input data vector $\mathbf{X}_n \in \mathbb{C}^{M_{\mathrm{L}}\times 1} = (x[n]\, x[n-1] \cdots x[n-M_{\mathrm{L}}+1])^T$. This method, however, requires storing an increased number of LUT entries, specifically $K^{M_{\mathrm{L}}+1}$, K being the number of LUT entries required in the simple case. An additional feature to add to a LUT model is an interpolation scheme [163], [189], which generally improves the approximation of the inverse of the PA's nonlinear characteristics by reducing the quantization error, while still maintaining a low number of LUT entries.

There are many other possible alternatives to modeling the nonlinear behavior

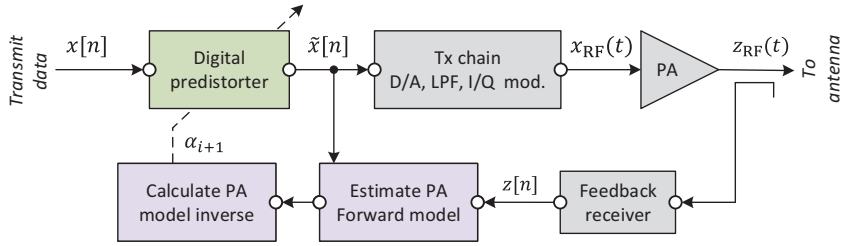
of a PA, which are not discussed within this thesis. In order to provide some insight, other example approaches could be artificial neural networks (ANN) [126], [167], [168], [186], [201], or the relatively new nonlinear auto-regressive moving average models (NARMA) [93], [96], [177], [195], [213]. The reader can find further details of these models in the provided articles and in references therein.

DPD Learning Architectures

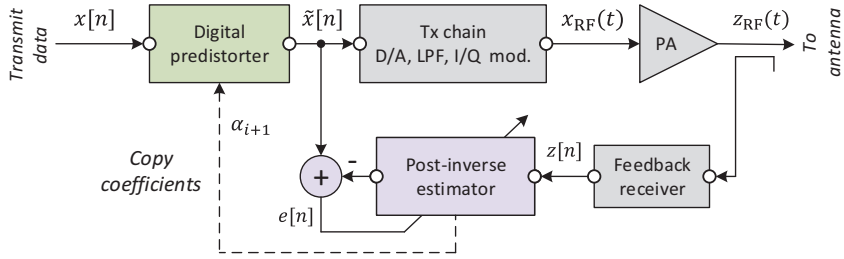
This subsection presents the main DPD architectures which are used for DPD coefficient estimation. These schemes are the direct learning architecture (DLA), indirect learning architecture (ILA), and closed-loop architecture, and are conceptually presented in Fig. 2.3, respectively.

First, the DLA is divided into a two-step process. In the first step, the PA direct model is estimated from the PA input and PA output signals, as depicted in Fig. 2.3a. In the second step, this estimate is directly inverted to be used as the DPD correction [7], [72], [135], [159], [177]. Examples of approaches for inverting the PA estimate are the pth-order inverse of nonlinear systems [165], [227], or the use of iterative inversion algorithms [7], [28], [72], [265]. One clear drawback of the DLA architecture is that the estimation of the PA inverse becomes more complicated when additional memory effects are considered in the DPD model, and inaccuracies in this process result in an accentuated performance degradation. This is the reason why this model is usually limited to memoryless systems. The DLA is said to be an open-loop technique, as the DPD function does not affect the coefficient estimation.

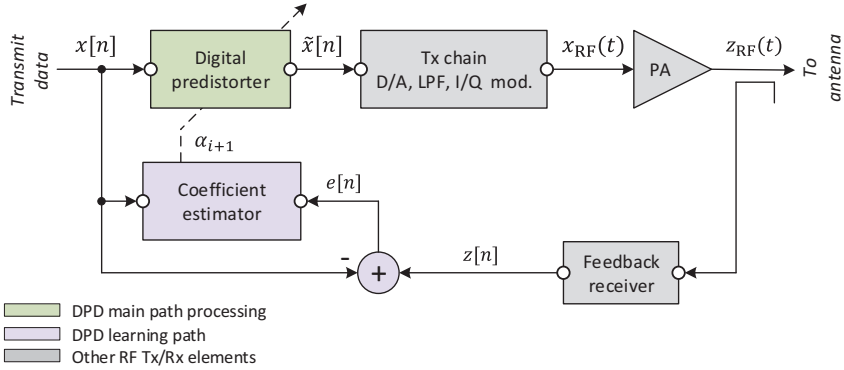
Second, the ILA is considered as one of the most widely used DPD schemes, as it is a very simple yet potent solution [22], [91], [177], [194]. In the ILA, the PA model post-inverse is extracted from the PA input and output signals, by comparing them and minimizing the generated error signal [19], [73], [83], [189], [249], [253] (i.e., if the error signal was zero, the PA input would equal the PA output), as depicted in Fig. 2.3b. In the digital post-distorter, the inverse PA nonlinearity is directly estimated. Such an inverse response is then directly copied to the pre-distorter, assuming that the pre- and post-estimations are equal. This approach avoids the inversion step which was required in the DLA, thus providing a very simple solution for linearizing nonlinear systems. However, it also has some drawbacks. In theory, the pre-inverse and post-inverse are not necessarily equal, when considering the exact inverse of a



(a) Direct learning architecture (DLA).



(b) Indirect learning architecture (ILA).



(c) Closed-loop learning architecture.

Figure 2.3 State-of-the-art DPD learning architectures most often utilized.

nonlinear system [184]. Also, the inverse function is undefined in saturated regions, as it may tend to infinity. This can be solved by just considering a specific threshold with respect to the input envelope, so that the DPD solution converges [177], [184]. The ILA is also an open-loop technique, with the DPD function not affecting the coefficient adaptation. Both open-loop DLA and ILA techniques are also referred to as self-tuning regulator architectures [177], [199].

Third, the closed-loop architecture estimates the DPD coefficients directly from the input system signal and PA output signals [38], [40], [189], [207], as depicted in Fig. 2.3c. This is done by generating the so called closed-loop error signal [P2, P4], which essentially contains the PA distortion. This error signal is then used to update the DPD coefficients, such that the PA output distortion is minimized [177], [184], [257]. This architecture is also known as a model reference adaptive system (MRAS) [35], [177], and forms a closed-loop technique, as the DPD function is inside the estimation block.

2.2.2 State-of-the-art in DPD and Linearization Techniques

This section presents a short summary regarding the state of the art in low complexity DPD techniques, as per the Author's knowledge up to date. The state-of-the-art summary begins with classical DPD approaches, which are applicable to any SISO device, and basically gather the most general and known predistorting solutions existing today. Then, a specific focus is given to low-complexity DPD approaches, where other alternative solutions focusing on reduced-complexity approaches are presented.

General DPD and Linearization Techniques

Digital predistortion and PA linearization techniques have been studied since the early 90s, resulting in many approaches which are generally considered by the scientific community as effective ways to increase the energy efficiency of TX PAs, while still transmitting high output power levels [8]. Some of these approaches are the Volterra-based series [38], [87], [91], [267] and its multiple subset models, such as the GMP [91], [193], [237], or MP [5], [91], [135], [237]. These methods usually provide robust and reliable DPD estimation, but also involve a substantial computational complexity, which poses a challenge for real-time DPD applications. These models are introduced at the beginning of this chapter as classical approaches, thus they are not reviewed in detail herein.

Reduced-complexity Cascaded Linearization Techniques

Cascaded nonlinear models are built upon serial connections of instantaneous

nonlinearities and LTI filters [71], [73], [91], [122], [132], [166], [255]. The main benefit compared to the linear-in-parameters models is that the model complexity in terms of number of coefficients is usually much smaller, although their estimation step is typically more complex. One recent example is provided by Dr. Scarpiniti and his research group, who have been recently studying cascaded structures applied to PA behavioral modeling [224]–[226]. In [225], they proposed the use of a Hammerstein structure in combination with a spline interpolation scheme to reduce the number of modeling parameters, compared to classical polynomial approaches. The work in [224] was equivalent, but proposed a Wiener-type structure instead. In [226], they further explored additional combinations of the Hammerstein-Wiener and Wiener-Hammerstein structures to enhance modeling capabilities, while also maintaining a low modeling complexity. Another example can be found in [260] by Dr. Younes et al., where the use of a three-box model consisting of a memoryless LUT, a memory polynomial, and an envelope memory polynomial (EMP), connected in parallel, was proposed. This model was basically an extension of the twin nonlinear two-box (TNTB) model, previously presented in [112], and it aimed at extending the cross-terms of the MP model with the addition of the EMP processing. Additionally, the nonlinear order in the MP model could be reduced as well, considering just a mild nonlinearity, as the LUT models the highly nonlinear behavior of the PA. This work was oriented to reduce the overall complexity in the predistorter. A third example is provided by Wu et al. in [255], where a parallel amplitude-to-amplitude modulation (AM/AM) and amplitude-to-phase modulation (AM/PM) spline-based Hammerstein architecture was configured for DPD applications. In this work, it was demonstrated how the number of parameters to adapt was reduced thanks to the cascaded structure. The coefficient adaptation was done with a least squares (LS) technique, which involved a quite heavy computational complexity in the learning stage.

Reduced-complexity LUT Linearization Techniques

LUTs can be used to implement arbitrary nonlinear functions, and they are often used in low-complexity DPD implementations [9], [62], [91], [105], [130], [189], [207], [264]. Typically, LUTs are used in memoryless DPD systems, to implement, for instance, the complex nonlinear gain function of the predistorter, or the AM/AM and AM/PM responses of the PA. LUTs can also be used as building blocks of other

nonlinear models to replace their polynomial computations. The accuracy of LUTs is generally proportional to the LUT size. However, the larger the LUT, the longer it will take to train all table entries, thus it is important to minimize the LUT size [46], [65]. To this end, Prof. Cavers [46] studied the optimum table spacing in LUT-based predistorting methods, by deriving an optimum companding function for the signal amplitude which minimizes the table size. Another approach to minimize the LUT size is *interpolation* of the output values. In the literature, linear interpolation [86], [189], quadratic interpolation [163], and spline interpolation [100], [224]–[226] have been applied to minimize the size of LUTs, with linear interpolation being the simplest technique out of these. These schemes allow reduction of the entry-size in the LUT, as well as speed up the LUT convergence, at the cost of a small increase in the processing complexity.

There have been many works that have addressed the memory addition to LUT-based predistorters. Nested LUTs are a typical solution [91], [264], but the LUT entries increase exponentially as memory terms are added. One example is the 2D-LUT proposed by Zhi-yong et al. in [264], where the power average of an arbitrary number of past samples was used to address the vertical and horizontal dimensions of the 2D-LUT. Another example was proposed by Dr. Jardin et al. in [130], who configured an additional codebook of filters after the LUT to add memory effects, addressed by the index of the LUT itself, while maintaining low processing complexity. This model can be seen as a Hammerstein-type architecture, but with an amplitude-dependent linear filtering part. In this work, the LUT does not have any interpolation, thus producing large LUT sizes, but still low compared to those of the nested LUT models. In [180], Dr. Ma et al. proposed an MP-LUT-based DPD model capable of compensating for the nonlinear and memory behavior of the PA. This structure avoided the use of nested LUTs, as the memory was intrinsically added by the MP branched structure. However, in the DPD learning update, all the LUTs appearing in each branch needed to be independently updated to provide an accurate memory correction.

Other Relevant Low-complexity State-of-the-art DPD Techniques

In this subsection, some other relevant works regarding low-complexity DPD are noted. A special mention is given to DPD hardware implementations, with emphasis on field programmable gate array (FPGA) and graphics processor unit (GPU)

works [36], [47], [105], [124], [137], [154], [261]. Commonly, the implementability of the DPD methods inherently implies a somewhat low complexity of the algorithms, as the resources in those devices are usually well constrained and limited. Good examples of hardware implementations can be found in [105], [161], [207]. In [207], Dr. Presti et al. proposed an FPGA implementation with the goal of maximizing the modulated output power and power-added efficiency while maintaining target ACLR numbers. Furthermore, in [105], Dr. Guan et al. presented an FPGA implementation of a DPD model based on dynamic deviation reduction (DDR) Volterra series. Finally, Dr. Li et al. presented in [161] an augmented parallel Hammerstein model implemented in a GPU, which was capable of compensating for the PA distortion, I/Q imbalance, and local oscillator (LO) leakage, inherently present in direct-conversion transmitters.

Finally, an additional short mention is given to sample-rate reduction methods. By oversampling the input signal, the aliasing issue is mitigated, as the new limit set by the Nyquist frequency accommodates also the signal frequencies produced by the nonlinear functions. However, using oversampling may lead to excessively high sample rates and processing complexities, due to the need for interpolation and decimation filters, which at the same time also increases the memory depth required in the DPD. Hence, there are several studies in the literature that focus on decreasing the required DPD sampling rates. Specifically, Dr. Ying et al. proposed in [172] an expansion of the real-valued Volterra series by introducing a set of band-limiting filter functions. By applying bandpass filtering corresponding to the band-limiting produced by the functions on the PA output, the linearization performance of the DPD is unaffected by the band-limiting scheme. A subsequent work was presented in [262], which showed good linearization performance even with low bandwidths for the feedback path if the PA output was bandpass filtered. In [234], the band-limited MP model was revised, making the cavity filter redundant, and a modified LS learning scheme was adopted. This work was further extended to an FPGA implementation in [250]. These methods focused on the band-limitation of both the main and learning paths, but it is also possible to just band-limit the feedback path, as described in [39], [72], [107], [171], [179] and references therein.

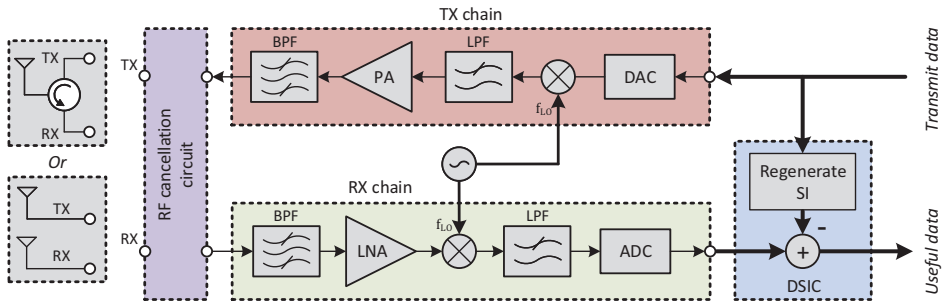


Figure 2.4 A complete direct-conversion IBFD architecture, showing both separate and shared TX/RX antenna configurations. The RF and digital cancellation stages are also shown.

2.3 Digital SI Cancellation for IBFD Transceivers

IBFD technology constitutes a recent discovery in the world of modern wireless communications, aiming to double the spectral efficiency by transmitting in the same frequency band and doing so at the same time. This technology can really boost the data rates in future wireless communication systems, but also poses several new challenges [30], [51], [53], [77], [145]. The most critical one is the self-induced SI signal, which is inevitably leaked from the TX node into the RX node. For proper IBFD operation, this problem needs to be dealt with, and hence it is also important to explore ways to deal with it by minimizing the involved computational complexity [81], [141]. The SI problem is discussed in depth in the the next subsection, and constitutes also one of the foci of this thesis.

2.3.1 Basics of Digital SI Cancellation

In order to build the theoretical background for studying digital SI cancellation, an example showing a complete IBFD architecture is depicted in Fig 2.4, showing also the RF and digital cancellation stages. This architecture can be configured with two different physical configurations: either separate TX and RX antennas in the TX/RX nodes, or a shared TX/RX antenna [15], [30], [140], [144]. With the former configuration, RF isolation can be obtained by strategically placing the antennas such that the TX leakage is minimized. One example is the use of a back-to-back re-

lay antenna as the one utilized for the experimental verification of this thesis [P3, P5]. With the latter, RF isolation can be achieved by using a duplexer or a circulator, which permits bidirectional communication over a single path [181], [205]. A circulator is capable of providing +30-40 dB of signal attenuation between the isolated ports, while presenting an insertion loss of just +0.3-1 dB for the desired signal path [205]. Another solution that allows for sharing a common TX/RX antenna is to use an electrical-balance duplexer (EBD), as the one used in [21]. An EBD builds upon a hybrid junction that connects the TX node, RX node, antenna port, and balancing impedance circuit port, and its TX/RX isolation is based on the accuracy of the impedance matching between the antenna and the impedance-balancing network [164], [182]. Additionally, an EBD can be easily implemented on a chip, and thus constitutes a good alternative for handheld devices. An EBD, however, is inherently narrow-band, and requires active tuning to track the antenna impedance at all times, thus it constitutes a power-driven component, in contrast to a circulator which is just a passive element.

Digital self-interference cancellation (DSIC) in IBFD is no more than a direct modeling problem, where the digital SI canceller aims at modeling the complete TX signal, from the TX digital-to-analog converter (DAC), all the way to the RX analog-to-digital converter (ADC). It then removes the generated replica from the complete received signal, ideally leaving only the desired signal contribution [53], [77], [142]. The most critical components in an IBFD device are the TX PA and the SI channel between the TX/RX antennas, or alternatively the circulator leakage. Since the TX PA typically exhibits deep nonlinear effects, stemming from the high saturation powers set to maximize its power efficiency, the overall SI modeling is also a nonlinear problem [12], [209], [217]. To tackle this issue, many different nonlinear models can be adopted [26], [121], [139], [142], as the ones already introduced when reviewing basic models for DPD. These models also constitute classical approaches, while novel and less complex solutions are derived within the next chapters of this thesis. The proposed models are applicable to an IBFD device having two separate antennas, shared antennas, or utilizing an EBD.

The overall received signal in an IBFD device contains the useful signal to be demodulated and its own SI, leaked from the TX chain of the transceiver. The overall scheme of the TX/RX of an IBFD transceiver is depicted in Fig. 2.5, showing all the involved signals. The baseband model of the received signal, $z(t)$, can be mathemat-

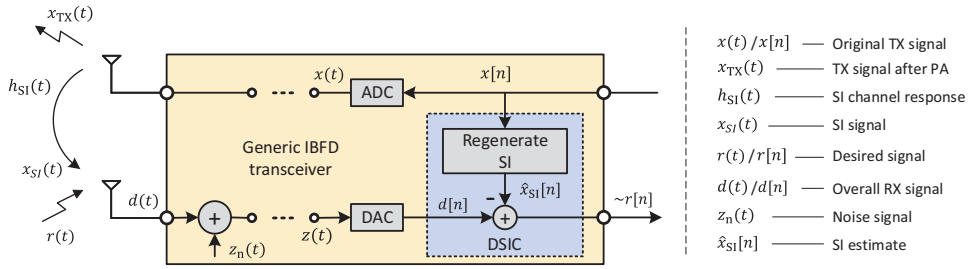


Figure 2.5 Conceptual illustration of the involved signals in a generic baseband IBFD transceiver, before and after the ADC/DAC.

ically expressed as

$$\begin{aligned} z(t) &= d(t) + z_{\mathrm{n}}(t) \\ &= r(t) + h_{\mathrm{SI}}(t) \star x_{\mathrm{TX}}(t) + z_{\mathrm{n}}(t) \\ &= r(t) + x_{\mathrm{SI}}(t) + z_{\mathrm{n}}(t), \end{aligned} \tag{2.11}$$

where $d(t)$ is the received signal, containing both the desired component (i.e., $r(t)$) and SI component (i.e., $x_{\mathrm{SI}}(t) = h_{\mathrm{SI}}(t) \star x_{\mathrm{TX}}(t)$), $z_{\mathrm{n}}(t)$ is random noise, $h_{\mathrm{SI}}(t)$ is the SI channel response, and $x_{\mathrm{TX}}(t)$ is the TX output signal. After having estimated the SI signal, the digital SI canceller removes it from the overall digitized received signal, $z[n]$, and thus the final expression of the received signal after the digital canceller is

$$\begin{aligned} z[n] &= d[n] - \hat{x}_{\mathrm{SI}}[n] + z_{\mathrm{n}}[n] \\ &= r[n] + x_{\mathrm{SI}}[n] - \hat{x}_{\mathrm{SI}}[n] + z_{\mathrm{n}}[n] \\ &= r[n] + e[n] + z_{\mathrm{n}}[n], \end{aligned} \tag{2.12}$$

where $\hat{x}_{\mathrm{SI}}[n]$ is the SI estimate, and the error signal $e[n]$ is caused by the imperfect DSIC performance. The SI estimate is typically generated by using signal models which take into account both the nonlinear response of the TX PA and the SI multipath channel. However, there exist other models that take into account further components appearing in the TX/RX chains. One particular example is the proposed spline-based Hammerstein-Wiener (SPHW) model presented in Section 3.5 and in [P7], which also considers the nonlinear response of the RX LNA, further enhancing the DSIC performance in selected scenarios.

2.3.2 State-of-the-art in Digital SI Cancellation

Transmitting and receiving at the same time and in the same frequency band is a type of communication that has been explored for already many years. Early deployments of radar technology date back to the 1940s, with the works in [27], [211], [233] being good examples of them, and [223] presenting a survey summarizing several post-war radar developments. The said works focused on STAR technology in continuous-wave and frequency-modulated radar, which transmit pulses that hit the targets to be detected, and are then bounced back to the radar system itself. Such a system receives the echoes while still uninterruptedly transmitting its own pulses to track or detect more targets. The main challenge of radar technology has always been the self-induced signal leaked into the receiver of the system itself, as a result of its own transmission. Obviously, the magnitude of this self-induced signal is significantly stronger than the echoes received from the targets, thus some kind of isolation procedure needed to be implemented. Early approaches typically used two separate TX and RX antennas, which provided certain isolation between the TX and RX system chains. This, on the other hand, generated also other challenges, as with two separate antennas additional fixed-target and moving-target interference arises [27]. Fixed-target interference is generated when the transmit signal is bounced back from a fixed object, such as a mountain or a building, hence having the same frequency as the transmit waveform. Moving-target interference is caused by a moving target bouncing the transmit waveform back with a different frequency, resulting from the Doppler effect [27]. These added problems, together with the own self-induced signal resulted in an overall low TX/RX isolation, causing early radar applications to suffer from reduced transmit powers, and hence an overall reduced detection distance range.

Since the early 90s, STAR technology has begun to be explored also in the context of radio communications. In fact, the aforementioned self-induced signal problem in early radar deployments is rather similar to the modern SI appearing in IBFD technology. Since any IBFD device must perform the SI cancellation to become commercially feasible, many results and developments can be taken from radar technology, and be also applied to IBFD. Ever since the aforementioned papers, many advances have been made regarding SI cancellation, and now it undoubtedly is a topic under intensive research by the scientific community, as the usage of IBFD technology is a

clear candidate for future wireless communications.

The vast majority of modern existing DSIC works consider the effect of the TX PA nonlinearity in combination with the self-interference channel, since they typically induce the strongest effect in the SI signal out of all the TX/RX chain components. Because of the nonlinear effect of the PA, SI cancellation techniques using purely linear cancellation approaches do not provide sufficient SI cancellation performance [30], [143], [162]. There are also many works in the literature which take into account further hardware imperfections, such as the TX/RX I/Q mismatch or the nonlinear effect of the RX LNA [138]. All of these works are reviewed in the following lines, especially focusing on time-domain digital SI cancellation, although SI cancellation has also been explored in the frequency domain [16], [53], [54].

The works presented in [20], [30], [79] focused on cancelling the SI signal under strong PA-induced nonlinear distortion. Specifically, the work in [30] considered an IBFD transceiver implemented with a circulator and hence a single antenna. In this system, the SI linear channel contribution was estimated with a linear system of equations that considered the transmitted packet preambles as known. The nonlinear TX PA contribution was estimated with a simple memoryless polynomial, thus assuming that no memory was generated by the PA. In [20], Dr. Anttila et al. proposed a parallel Hammerstein approach to model both the effects of the nonlinear TX PA and the SI channel. In general, the parallel Hammerstein model has been reported to be a versatile yet reliable solution for behavioral modeling of PAs [75], [127], and it was seen to be a robust approach also in the context of IBFD. The proposed technique made it possible to transmit higher levels of TX power, while the TX PA nonlinearity and the SI channel were compensated by the proposed DSIC. Finally, the work by Emara et al. in [79] proposed a GPU implementation of a IBFD transceiver with a single antenna and a circulator. In this work, both of an analog and digital cancellation stages were used. The authors utilized a classical MP approach to model the PA-induced nonlinearity and channel response, with a transversal recursive least squares (RLS) to adapt the MP coefficients, which was based on [80]. The total amount of achieved SI cancellation was +90 dB, out of which the DSIC achieved +36 dB.

There have been also some works which merge DPD with IBFD in order to reduce the complexity of the SI cancellation stage. In theory, DPD reduces the overall nonlinear distortion induced by the TX PA, thus a simpler DSIC can be used in the

resulting transmitter. Examples of these techniques can be found in [23], [102], for instance. Overall, these works proposed the use of a memory polynomial to predistort the signal prior to the SI cancellation stages. The DPD model reduced the side lobes of the overall transmit waveform, and the DSICs were able to achieve an enhanced cancellation performance when compared to the case where no DPD was applied.

In [12], [25], [32], [147]–[149], [152], [219], the reader can find examples of DSICs which aim to cancel further hardware impairments, such as I/Q imbalance or LNA distortion. The I/Q imbalance problem specially happens in direct conversion architectures (or homodyne architectures), where the signal is upconverted and downconverted directly from/to baseband to/from RF [30], [77], [145]. In this type of architecture, the image signal corresponds also to the desired signal itself, and the non-perfect self-image rejection effect is found as a transformation of the original signal constellation. Hence, the image rejection requirements are not very high, with +30-40 dB being sufficient in most practical cases [53], [143]. Nevertheless, other I/Q compensation techniques can be applied to further mitigate the self-induced image effect [18], [20]. In general, the DSIC proposals which also aim at cancelling the I/Q imbalance effect are able to enhance the cancellation performance, but also have the cost of increased computational complexity. On the one hand, the techniques presented in [148], [219] constituted advanced DSICs aiming at compensating for the I/Q impairments produced in the TX and RX mixing stages. These techniques utilized the widely linear cancellation principle to take this into account. For the coefficient estimation, LS was used in [148], while an adaptive filtering technique which minimized the error signal between the original and regenerated SI signal was used in [219] to facilitate continuous tracking. On the other hand, the techniques studied in [12], [147] focused on compensating for the nonlinear LNA effect. Both techniques utilized polynomial functions to model the RX nonlinear distortion, as also presented in [103], and used a linear LS algorithm to estimate the model coefficients. In [147], a design of an improved compact relay antenna with a monoplane rotation of +45° was also introduced and shown to improve the RF isolation to a minimum of +55 dB. The work presented by M.Sc. Kurzo et al. in [152] presented an ANN scheme for cancelling the SI signal, alongside with its corresponding main-path FPGA and application-specific integrated circuit (ASIC) implementation. The SI cancellation was divided into two stages: a standard linear cancellation procedure

to regenerate the linear SI component, and a two-layer ANN to reconstruct the nonlinear part of the SI. With the hardware implementation, it was demonstrated that the ANN achieved higher throughput with an overall lower resource usage. A similar work was presented in [24].

It is finally noted that further good examples of DSIC prototype implementations can be found in [14], [49], [78], [81], [110], [128], [160], [231], [235], [254], and references therein.

3 SPLINE-BASED CASCADED MODELS

THIS chapter presents different injection-based and spline-interpolated cascaded models to be used for either digital predistortion (DPD) or in-band full-duplex (IBFD) applications, among others. In particular, Hammerstein, Wiener, Wiener-Hammerstein, and Hammerstein-Wiener architectures are presented, alongside their corresponding gradient-based parameter estimation learning stages.

The contents of this chapter are based on the discoveries found in the publications [P1, P3, P5, P7], as well as on the works presented in [22], [41], [64], [71], [130], [163], [189], [206], [221], [224]–[226], [241], [255].

3.1 Review of B-spline Interpolation Theory

This section presents the basics regarding B-spline interpolation theory. B-spline interpolation is first presented in the context of real-valued data, and is then extended to the complex-valued case to be also used with complex-valued I/Q signals.

3.1.1 Real-valued B-spline Interpolation

Spline interpolation schemes build on piece-wise polynomials, which interpolate through an arbitrary set of points generally known in this context as *control points*, having smoothness and continuity constrains imposed on the connecting points [64], [218]. Taking such piece-wise approaches poses several advantages, compared to classical polynomial models where a high-order expression is utilized to model the whole input data range. The main advantage is that simpler and lower-order functions can be taken per individual region in the model [9], [183], [206]. The use of lower-order functions essentially translates to reduced associated complexities [189], [225], [255], as multiplications rapidly increase with the polynomial order. It also allows for better conditioning of the estimation problem, obtaining a lower condition number in

the regression matrix [256] that avoids the need of prewhitening or orthogonalization [P1]. Also, a different spline order and memory parametrization can be taken per individual model region, depending on the desired modeling accuracy in each.

In order to build this particular piece-wise approach, let us define a set of *knots* which divide the input data range into exactly N^{SP} regions, as $\mathbf{T} = \{t_0, t_1, \cdots, t_{N^{\text{SP}}}\}$, mapping to regions $i = \{0, 1, \cdots, N^{\text{SP}} - 1\}$. Additionally, the *knot* sequence is constrained to be non-decreasing, i.e., $t_0 < t_1 < \cdots < t_{N^{\text{SP}}}$ [206]. In the context of both DPD and IBFD, an equi-spaced spline interpolation scheme is assumed, which allows for simpler input-output relations and simpler update equations in posterior models [224], [225]. Hence, the region width is defined as $\Delta = t_{i+1} - t_i$, $\Delta > 0$, and is considered invariant within regions.

The construction of a B-spline interpolation scheme can be generalized to any number of regions N^{SP}, any number of *control points*, and any spline polynomial degree P. The B-spline curve fitted in each segment, $[t_i, t_{i+1})$, is an affine combination of $P+1$ curves that form the overall polynomial in the corresponding region i. The Pth-degree spline basis function characterizes each of these curves, and it is given by the De Boor recursion [64], defined as

$$N_i^P(u) = \frac{u - t_i}{t_{i+P} - t_i} N_i^{P-1}(u) + \frac{t_{i+P+1} - u}{t_{i+P+1} - t_{i+1}} N_{i+1}^{P-1}(u), \tag{3.1}$$

where the basic 0th-order basis function, $N_i^0(u)$ can be expressed as

$$N_i^0(u) = \begin{cases} 1 & \text{if } t_i \leq u < t_{i+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

It is noted that each spline basis function, $N_i^P(u)$, is non-zero exclusively in the interval $[t_i, t_{i+P+1})$. Additionally, the spline basis functions $N_i^P(u)$ in different regions are just shifted versions of each other, i.e., $N_i^P(u) = N_0^P(u - i)$. The B-spline basis functions $N_i^0(u)$, $N_i^1(u)$, and $N_i^2(u)$, corresponding to P degrees 0, 1, and 2, respectively, are showed conceptually in Fig. 3.1. The complete B-spline curve in an arbitrary region i can be obtained by combining the $P+1$ spline blending functions

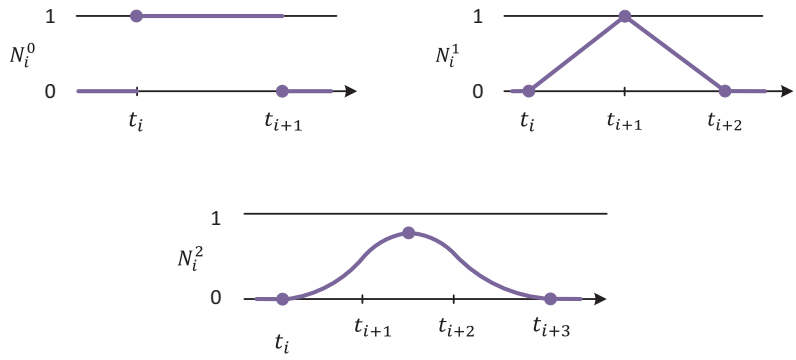


Figure 3.1 B-spline basis functions $N_i^0(u)$, $N_i^1(u)$, and $N_i^2(u)$, of order 0, 1, and 2, respectively. Equi-spaced *knots* are assumed.

as

$$\gamma(u,i) = \sum_{n=i-P}^{i} N_n^P(u) q_n, \tag{3.3}$$

where q_n is the control point which is weighing each spline curve.

A more generic expression of $\gamma(u,i)$ can be obtained by formulating it with a two-vector inner product, and by further substituting each spline basis matrix shown in (3.8). Its final form, for each spline segment, reads

$$\gamma(u,i) = \begin{bmatrix} u^P & u^{P-1} \cdots & 1 \end{bmatrix} \mathbf{C}_P \begin{bmatrix} q_{i-P} \\ q_{i-P+1} \\ \vdots \\ q_i \end{bmatrix}, \tag{3.4}$$

where $\mathbf{C}_P \in \mathbb{R}^{(P+1)\times(P+1)}$ is the B-spline basis matrix, which is directly obtained from the De Boor recursion and depends on the curve order P and region width Δ. Examples of this matrix for B-spline orders $P = 1, \cdots, 3$ are shown in (3.8). By combining all spline segments, the formula in (3.4) can be further generalized to describe the complete spline curve. Thus, it can be rewritten, with matrix notation,

$$\mathbf{C}_1 = \begin{bmatrix} \frac{-1}{\Delta} & \frac{1}{\Delta} \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{C}_2 = \frac{1}{2}\begin{bmatrix} \frac{1}{\Delta^2} & \frac{-1}{\Delta^2} & \frac{1}{\Delta^2} \\ \frac{-2}{\Delta} & \frac{2}{\Delta} & 0 \\ 1 & 1 & 0 \end{bmatrix}, \qquad \mathbf{C}_3 = \frac{1}{6}\begin{bmatrix} \frac{-1}{\Delta^3} & \frac{3}{\Delta^3} & \frac{-3}{\Delta^3} & \frac{1}{\Delta^3} \\ \frac{3}{\Delta^2} & \frac{-6}{\Delta^2} & \frac{3}{\Delta^2} & 0 \\ \frac{-3}{\Delta} & 0 & \frac{3}{\Delta} & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \tag{3.8}$$

as

$$\gamma\prime(u,i) = \mathbf{\Psi}^T \mathbf{q}, \tag{3.5}$$

where $\mathbf{q} \in \mathbb{R}^{Q^{\mathrm{SP}} \times 1}$ is now the complete set of control points. The total number of control points, for N^{SP} regions, is $Q^{\mathrm{SP}} = N^{\mathrm{SP}} + P$. The column $\mathbf{\Psi}$ and the vector $\mathbf{u}$ (directly extracted from (3.4)), read

$$\mathbf{\Psi} \in \mathbb{R}^{Q^{\mathrm{SP}} \times 1} = \begin{bmatrix} 0 & \cdots & 0 & \mathbf{u}^T \mathbf{C}_P & 0 & \cdots & 0 \end{bmatrix}^T, \tag{3.6}$$

$$\mathbf{u} \in \mathbb{R}^{(P+1)\times 1} = \begin{bmatrix} u^P & u^{P-1} \cdots & 1 \end{bmatrix}^T, \tag{3.7}$$

where the term $\mathbf{u}^T \mathbf{C}_P$ is indexed from the ith to the $(i+P)$th position in $\mathbf{\Psi}$, such that only the corresponding control points are selected for the interpolation. Note that, depending on the spline order P, the interpolation scheme will be characterized as:

- $P = 0$: Nearest neighbor interpolation.
- $P = 1$: Linear interpolation.
- $P \geq 2$: Higher-order interpolation (e.g., quadratic, cubic, etc.).

For a more detailed explanation about real-valued spline interpolation theory, refer to [64], [206], or [P2].

3.1.2 Complex-valued B-spline Interpolation

Traditionally, spline theory has been applied to real-valued systems [224]–[226]. However, in the context of radio communications the processing is based upon complex-valued I/Q data, thus spline theory needs to be extended to the complex domain. Let's begin with a simple memoryless baseband model of a bandpass nonlinear

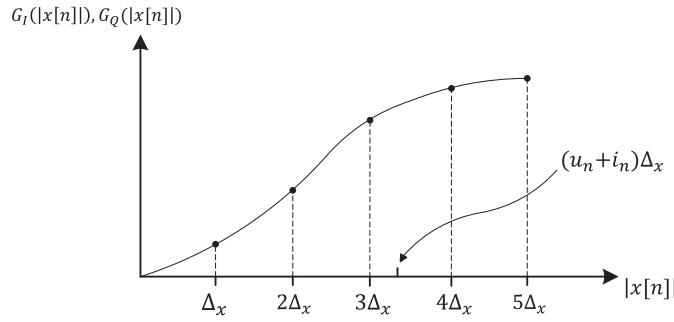


Figure 3.2 Conceptual illustration of the B-spline region partitioning of $G_{\mathrm{I}}(|x[n]|$ and $G_{\mathrm{Q}}(|x[n]|$ with respect to $|x[n]|$. Note that, in this particular example, five regions are considered, $i_n = 3$, and u_n takes some value in the interval $[0, \Delta_x)$.

device (e.g., a PA), with an input-output response expressed as [P1-P3, P5]

$$y[n] = x[n]G(|x[n]|) = x[n]\Big(G_{\mathrm{I}}(|x[n]|) + jG_{\mathrm{Q}}(|x[n]|)\Big), \tag{3.9}$$

where $x[n]$ is the input signal, and $G(|x[n]|)$ is the complex-valued nonlinear gain of the device, only depending on its amplitude and not its phase. The complex spline interpolation can be built from (3.9), considering $|x[n]|$ and $G(|x[n]|)$ as the spline input and spline output signals, respectively. Since the spline gain is complex-valued, two different splines modeling the I and Q data branches are considered. To characterize the interpolation scheme, two local parameters which depend only on the magnitude of the input signal are defined, named the region index, i_n, and the abscissa value, u_n. These are defined as

$$i_n = \left\lfloor \frac{|x[n]|}{\Delta_x} \right\rfloor + 1, \tag{3.10}$$

$$u_n = \frac{|x[n]|}{\Delta_x} - (i_n - 1). \tag{3.11}$$

The region index indicates the region in which $|x[n]|$ is located at time instant n, and the abscissa corresponds to the normalized value within the region i_n ($u_n \in [0, \Delta_x)$). These parameters are conceptually illustrated in Fig. 3.2, assuming in this example 5 model regions, $i_n = 3$, and $u_n \in [0, \Delta_x)$.

According to the real-valued theory presented in the previous section, the outputs

of the two independent splines modeling the I and Q branches can be expressed as

$$G_{\mathrm{I}}(|x[n]|) = \boldsymbol{\Psi}_n^T \mathbf{q}^{\mathrm{re}}, \tag{3.12}$$

$$G_{\mathrm{Q}}(|x[n]|) = \boldsymbol{\Psi}_n^T \mathbf{q}^{\mathrm{im}}, \tag{3.13}$$

where $\mathbf{q}^{\mathrm{re}} \in \mathbb{R}^{Q^{\mathrm{SP}}\times 1}$ and $\mathbf{q}^{\mathrm{im}} \in \mathbb{R}^{Q^{\mathrm{SP}}\times 1}$ are real-valued sets containing the control points for each I and Q branches, and $\boldsymbol{\Psi}_n$ is defined in (3.6). Substituting these expressions in (3.9), the combined complex-valued spline output can be expressed as

$$\begin{aligned} y[n] &= x[n]\Big(G_{\mathrm{I}}(|x[n]|) + jG_{\mathrm{Q}}(|x[n]|)\Big) \\ &= x[n]\boldsymbol{\Psi}_n^T\Big(\mathbf{q}^{\mathrm{re}} + j\mathbf{q}^{\mathrm{im}}\Big) \\ &= x[n]\boldsymbol{\Psi}_n^T\mathbf{q}, \end{aligned} \tag{3.14}$$

where $\mathbf{q} \in \mathbb{C}^{Q^{\mathrm{SP}}\times 1} = [q_0\ q_1\ \cdots\ q_{Q^{\mathrm{SP}}-1}]^T$ combines the control points of the I and Q branches into a complex-value set. Note that $\mathbf{q}$ can be seen as an LUT, the control points being the table entries. This expression constitutes the final complex-valued B-spline interpolation output.

3.1.3 Injection-based Complex-valued B-spline Interpolation

Once the complex-valued nonlinear output of the spline interpolation scheme is derived, it can be modified to be expressed as a deviation from the unit gain. This is referred to as an injection-based scheme [3], which will allow for certain benefits, explained in following lines. With such a formulation, (3.9) can be rewritten as

$$y[n] = x[n]\Big(1 + G_{\mathrm{I}}(|x[n]|) + jG_{\mathrm{Q}}(|x[n]|)\Big), \tag{3.15}$$

and the modified B-spline output can be expressed instead as

$$\begin{aligned} y[n] &= x[n]\Big(1 + \boldsymbol{\Psi}_n^T\mathbf{q}^{\mathrm{re}} + j\boldsymbol{\Psi}_n^T\mathbf{q}^{\mathrm{im}}\Big) \\ &= x[n] + x[n]\boldsymbol{\Psi}_n^T\mathbf{q} \\ &= x[n]\boldsymbol{\Psi}_n^T(\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}), \end{aligned} \tag{3.16}$$

where $\mathbf{1}_{Q^{\text{SP}}} \in \mathbb{R}^{Q^{\text{SP}} \times 1}$ denotes a column of all ones, and the partition of the unity property of the B-splines [64], [206], i.e., $\boldsymbol{\Psi}_n^T \mathbf{1}_{Q^{\text{SP}}} = 1$, was used to factorize the last expression. This expression constitutes the final complex-valued injection-based spline interpolated output which will be used in the course of this thesis.

Applying such formulation will have several benefits in the context of cascaded structures and also LUT systems, as demonstrated in [P1, P2] . Such benefits can be summarized as follows:

- Gain ambiguities between elementary LTI and nonlinear blocks in cascaded models can be effectively removed, which is relevant in posterior derivations of cascaded structures.
- The number of required bits to quantize the vector $\mathbf{q}$ in a fixed-point implementation is reduced, as its dynamic range is decreased. This effect was demonstrated in [P2, P3] by measuring the LUT bit quantization vs modeling performance in a modified Saleh model [197], [230].
- If the control point vector $\mathbf{q}$ is initialized as a zero vector (i.e., $\mathbf{0}_{Q^{\text{SP}}}$), the spline output signal is equal to the input signal, i.e., $y[n] = x[n]$.

The complete processing scheme of the injection-based spline interpolation scheme is conceptually presented in Fig. 3.3. Such a scheme will be used as a building block for the injection-based spline-interpolated scheme in the following cascaded structures. Finally, it is noted that a generic nomenclature is adopted within this section, where the input signal is named as $x[n]$, the intermediate model signals are denoted as $s[n]$ and $l[n]$ – according to the considered cascaded structure – and the output model signal is referred to as $y[n]$. It is also worth mentioning that in a DPD application, the signal $y[n]$ would correspond to the signal after the predistortion stage, named $\tilde{x}[n]$ in Section 2.2.1. Alternatively, in an IBFD application, the output signal $y[n]$ would correspond to the SI estimate, denoted as $\hat{x}_{\text{SI}}[n]$ in Section 2.3.1.

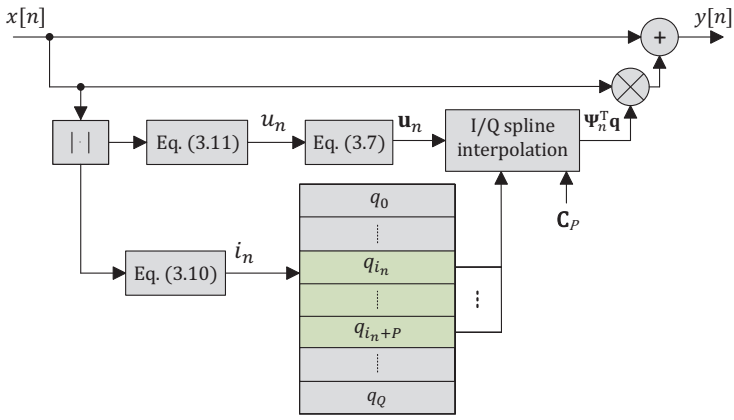


Figure 3.3 Complete scheme of the complex-valued injection-based spline interpolation scheme, utilized to model the nonlinear blocks in subsequent sections.

3.2 Spline-based Hammerstein Model

This section overviews the developed spline-based Hammerstein model, also in combination with the aforementioned injection-based scheme, as presented in publications [P1, P3, P5]. A Hammerstein-type system incorporates first a nonlinearity and second a linear block, which aim at modeling an unknown system whose structure is alike [71], [92], [166], [225], [255]. In this particular example, the nonlinear block is modeled with the aforementioned injection-based spline interpolation scheme, and the linear block is implemented as an LTI FIR filter. For the rest of the thesis, this model is referred to as SPH.

In the context of DPD, the nonlinear block aims at estimating the PA-induced nonlinearity, and the LTI filter aims at compensating for its memory [73]. In the context of IBFD, the nonlinearity only models the TX PA, while the FIR filter accounts for the effects of the SI multipath channel [148]. Further enhancements of the SPH model incorporating an extra block compensating for the inherent TX mixer I/Q imbalance can be found in [21], in the context of IBFD.

The complete architecture of the cascaded SPH model is presented in Fig. 3.4. In accordance to the nomenclature utilized in the figure, let us denote the input and output model signals as $x[n]$ and $y[n]$, and the output signal of the nonlinear block as $s[n]$. Following the B-spline interpolation processing described in (3.16), and

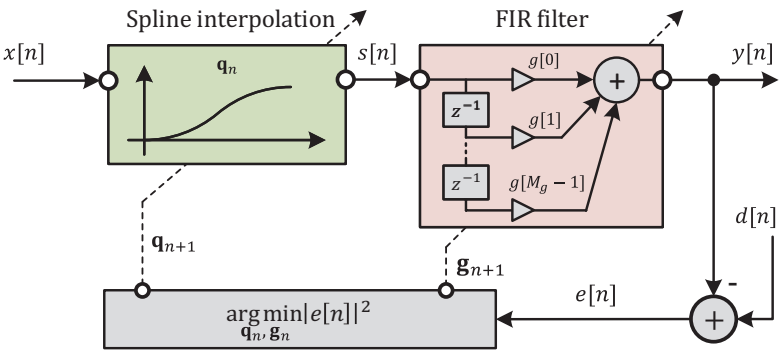


Figure 3.4 Complete architecture of the complex cascaded SPH structure. The *spline interpolation block* comprises the processing scheme presented in Fig. 3.3.

applying a classical convolution to the resulting signal, the intermediate and output signals of the SPH model can be written as

$$\begin{aligned} s[n] &= x[n]\boldsymbol{\Psi}_n^T(\mathbf{1}_Q + \mathbf{q}_n), \\ y[n] &= \mathbf{g}_n^T \mathbf{s}_n, \end{aligned} \tag{3.17}$$

where the subindex n indicates the time-dependency of the parameters, $\mathbf{g}_n \in \mathbb{C}^{M_g \times 1} = [g[0]\, g[1] \cdots g[M_g{-}1]]^T$, $\mathbf{s}_n \in \mathbb{C}^{M_g \times 1} = [s[n{+}M_{g,\text{pre}}] \cdots s[n] \cdots s[n{-}M_{g,\text{post}}]]^T$ is the signal regression of $s[n]$, and $M_g = M_{g,\text{pre}} + M_{g,\text{post}} + 1$ denotes the total number of taps of the linear FIR filter, $M_{g,\text{pre}}$ and $M_{g,\text{post}}$ being the pre- and post-cursor memory taps. For simplicity, $M_{g,\text{pre}} = 0$ is considered for the rest of the thesis, which significantly simplifies the chosen notation.

3.2.1 Learning Equations

Once having derived the input-output relation of the SPH model, two different learning rules estimating the spline control points, $\mathbf{q}_n$, and the FIR filter, $\mathbf{g}_n$, can be derived to make the SPH model adaptive. To this end, an error signal $e[n]$ is first defined as

$$e[n] = d[n] - y[n], \tag{3.18}$$

where $d[n]$ is the observed or desired signal containing the response of the system. The problem now lies in minimizing the error signal with respect to both parameters to adapt, namely $\mathbf{q}_n$ and $\mathbf{g}_n$. There are several ways of minimizing a function, but the one used for this work is the gradient-descent solution [41], [116], [192], [252], where the parameters are learned by following the steepest negative direction of the gradient of a *cost function*, until a local minimum in the mean squared error (MSE) is reached [115]. The cost function to be minimized can be defined as the instantaneous squared error. Such a cost function depends on both parameters to estimate, and can be formally defined as

$$J(\mathbf{q}_n, \mathbf{g}_n) = |e[n]|^2 = e[n]e^*[n]. \tag{3.19}$$

Taking the gradient-descent approach, the general learning rules for adapting the control points and the FIR filter read

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_\mathrm{q} \nabla_{\mathbf{q}_n} J(\mathbf{q}_n, \mathbf{g}_n), \tag{3.20}$$

$$\mathbf{g}_{n+1} = \mathbf{g}_n - \mu_\mathrm{g} \nabla_{\mathbf{g}_n} J(\mathbf{q}_n, \mathbf{g}_n), \tag{3.21}$$

where μ_q and μ_g are the learning rates for the control points and filter coefficients, respectively, and ∇_x represents the gradient operator, evaluated with respect to x. Invoking elementary differentiation rules, the resulting learning rules read

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_\mathrm{q} e[n] \boldsymbol{\Sigma}_n^T \mathbf{X}_n^* \mathbf{g}_n^*, \tag{3.22}$$

$$\mathbf{g}_{n+1} = \mathbf{g}_n + \mu_\mathrm{g} e[n] \mathbf{s}_n^*, \tag{3.23}$$

where $\mathbf{X}_n \in \mathbb{C}^{M_\mathrm{g} \times M_\mathrm{g}}$ gathers the signal regression of $x[n]$ in its main diagonal, and $\boldsymbol{\Sigma}_n \in \mathbb{R}^{M_\mathrm{g} \times Q^\mathrm{SP}} = [\boldsymbol{\Psi}_n \ \boldsymbol{\Psi}_{n-1} \ \cdots \ \boldsymbol{\Psi}_{n-M_\mathrm{g}+1}]^T$. Moreover, in the learning rule for the spline control points, it is assumed that the rate of change of the control point vector over the span of the FIR filter is negligible (i.e., $\mathbf{q}_n \approx \mathbf{q}_{n+M_\mathrm{g}}$). This assumption is reasonable since the learning rate μ_q is small. Detailed derivations can be found in [P1].

It is finally noted that both parameters in the SPH model are independent, thus they can be freely adapted depending on the requirements on the application hosting the algorithm [76], [116]. For instance, when applying such a method in an IBFD device, the nonlinear response estimate, $\mathbf{q}_n$, once converged, can usually remain con-

stant, since the behavior of the PA is not expected to change noticeably. Meanwhile, the SI channel coefficients, $\mathbf{w}_n$, can be continuously estimated, stemming from the ever-varying environmental multi-path effects.

3.2.2 Complexity Analysis

In this section, the computational complexity of the SPH model is analyzed. The complexity is divided into two stages, model identification and coefficient update, and is presented in terms of real multiplications per sample. This metric is deliberately chosen, as multiplications constitute, in general, the most resource-demanding operation in digital signal processor (DPS) devices [237], while additions cost considerably less [105]. Multiplications are a key parameter, as resources in HW devices, such as FPGAs, are usually limited. Herein, it is assumed that one complex-complex multiplication can be obtained by four real multiplications, and one complex-real multiplication can be executed with two real multiplications [99]. Additionally, the square roots coming from the absolute value operator are left indicated as $sqrt$. One simple option to calculate the absolute value without the need for square roots is to use the alpha-max beta-min algorithm [88], [89], as utilized in [P1]. The complexity of the SPH and posterior models calculated in terms of floating point operations (FLOPS) can be found in [P1,P3].

The complexity of the SPH model is calculated following the exact processing steps taken in the previous section, and is presented in Table 3.1, as a function of the SPH modeling parameters. It is noted that the presented complexity expressions basically represent an upper bound, since multiplications by any power of 2 or 1/2 do not represent any added complexity in the final system, as they can be easily implemented with arbitrary bit shifts. An additional numerical example is presented also in Table 3.1, where the model parametrization has been chosen to be $P = 3$, $M_{\mathrm{g}} = 10$, and $\tau^{\mathrm{SP}} = 5$. This is a similar parametrization to the one taken for the verification of the DPD and IBFD applications.

At this point, it is important to note the relatively high complexity of the term $\mathbf{\Sigma}_n^T \mathbf{X}_n^* \mathbf{g}_n^*$ in the learning update of $\mathbf{q}_{n+1}$. To ease the update, a complexity-reduction method was proposed in [P1, P5]. In this method, the temporal dimension (i.e., rows) of $\mathbf{\Sigma}_n^T$ is truncated to contain only a reduced time span, which maps to the largest filter taps in $\mathbf{g}_n$. These taps usually correspond to the nearest past samples,

thus a span containing only τ^{SP} taps in $\mathbf{\Sigma}_n^T$ can be selected instead. Note that if $\tau^{\mathrm{SP}} = M_{\mathrm{g}}$, no complexity reduction would be adopted. As demonstrated in [P1, P5], this simplification has a minimal effect on the performance of the SPH model, while greatly reducing the required complexity of the learning algorithm.

Table 3.1 Computational complexity required to execute the model calculation and coefficient update of the SPH identification algorithm, in terms of real multiplications per sample. Also a numerical example is presented, where $P = 3$, $M_{\mathrm{g}} = 10$, and $\tau^{\mathrm{SP}} = 5$.

Operation		Symbolic real mul./sample	Numeric real mul./sample
Model calculation	$s[n]$	$P^2+4P+8+sqrt$	29
	$y[n]$	$4M_{\mathrm{g}}$	40
	Total	$P^2+4P+4M_{\mathrm{g}}+8+sqrt$	69
Coefficient update	$\mathbf{q}_{n+1}$	$4P+2P\tau^{\mathrm{SP}}+6\tau^{\mathrm{SP}}+6$	78
	$\mathbf{g}_{n+1}$	$4M_{\mathrm{g}}+2$	42
	Total	$4P+2P\tau^{\mathrm{SP}}+6\tau^{\mathrm{SP}}+4M_{\mathrm{g}}+8$	120
Total per iteration		$P^2+8P+2P\tau^{\mathrm{SP}}+6\tau^{\mathrm{SP}}+8M_{\mathrm{g}}+16+sqrt$	189

3.3 Spline-based Wiener Model

The spline-based Wiener model, denoted herein as SPW, is next developed, as presented in [P3]. The structure of this model is similar to the previous SPH, but the order of the LTI and nonlinear elementary blocks is now inverted. The resulting structure contains first the linear FIR filter which is followed by the spline-based nonlinear block [92], [101], [122], [224]. This architecture is conceptually presented in Fig. 3.5, which also shows the parameter learning stage.

Let us denote the input and output model signals as $x[n]$ and $y[n]$, and the intermediate signal after the FIR filter as $s[n]$. Their expressions, after applying a classical convolution for the FIR filter and the spline-interpolated scheme for the nonlinear block, can be expressed as

$$s[n] = \mathbf{w}_n^T \mathbf{x}_n, \tag{3.24}$$

$$y[n] = s[n]\mathbf{\Psi}_n^T(\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_n), \tag{3.25}$$

where $\mathbf{w}_n \in \mathbb{C}^{M_{\mathrm{w}}\times 1} = [w[0]\ w[1]\ \cdots\ w[M_{\mathrm{w}}-1]]^T$, and $\mathbf{x}_n \in \mathbb{C}^{M_{\mathrm{w}}\times 1} = [x[n]\ x[n-1]\ \cdots\ x[n-M_{\mathrm{w}}+1]]^T$ is the signal regression of $x[n]$.

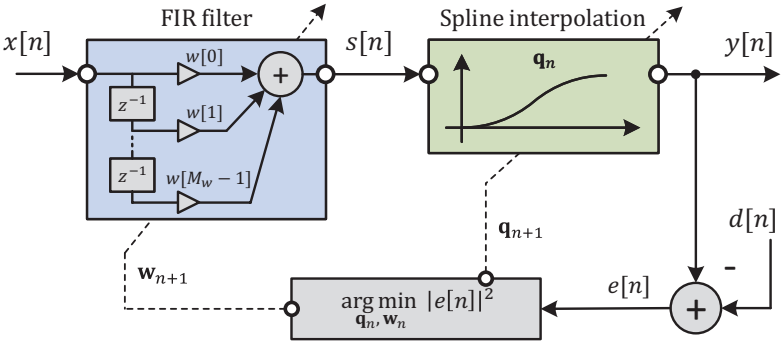


Figure 3.5 Complete architecture of the complex cascaded SPW structure. The *spline interpolation block* comprises the processing scheme presented in Fig. 3.3.

3.3.1 Learning Equations

The derivation of the SPW learning rules is done in a similar manner as the one adopted in the SPH model. The SPW associated error signal can be defined as in (3.18), and the new cost function can be expressed as in (3.19), but changing the filter with $\mathbf{w}_n$ instead. The same gradient-descent solution is adopted to estimate the quantities of $\mathbf{q}_n$ and $\mathbf{w}_n$. Applying such a formulation gives the following expressions for the learning rules

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_{\mathrm{q}} \nabla_{\mathbf{q}_n} J(\mathbf{q}_n, \mathbf{w}_n), \tag{3.26}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{\mathrm{w}} \nabla_{\mathbf{w}_n} J(\mathbf{q}_n, \mathbf{w}_n), \tag{3.27}$$

where μ_{w} is the learning rate for the preceding FIR filter. Further developing the expressions for both parameters, the final learning rules read

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_{\mathrm{q}} e[n] s^*[n] \boldsymbol{\Psi}_n, \tag{3.28}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_{\mathrm{w}} \mathbf{x}_n^* \Big(e[n] \boldsymbol{\Psi}_n^T (\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_n^*) + \frac{2 s[n] \mathrm{Re}\{e[n] \dot{y}[n]\}}{\Delta_s |s[n]|} \Big), \tag{3.29}$$

where $\dot{y}[n] = s^*[n] \dot{\mathbf{u}}_n^T \mathbf{C}_P (\mathbf{1} + \mathbf{q}_{i,n}^*)$, the column $\dot{\mathbf{u}}_n \in \mathbb{R}^{(P+1)\times 1} = [P u_n^{P-1} \; (P-1) u_n^{P-2} \; \cdots \; 1 \; 0]^T$ represents the derivative of the abscissa vector in (3.7), and $\mathbf{q}_{i,n} \in \mathbb{C}^{(P+1)\times 1} = [q_{i,n} \; q_{i+1,n} \; \cdots \; q_{i+P,n}]^T$ are the control points selected for the spline

interpolation. It is finally noted that these learning equations are independent, and can be updated freely as required by the final application hosting the algorithm.

3.3.2 Complexity Analysis

The computational complexity of the SPW model is divided into two stages, model identification and coefficient update, and is detailed in terms of real multiplications per sample. The same assumptions taken in Section 3.2.2 are adopted. Additionally in this case, the number of required divisions is left indicated as div.

The complexity of the SPW model follows the processing steps derived in the previous section, and is presented in Table 3.2, as a function of the SPW modeling parameters. Again, these expressions constitute an upper bound for the complexity, as many multiplications with powers of 2 or 1/2 are free in DSP implementations. Also, an additional numerical example is presented within the table, where the model parametrization has been chosen to be $P = 3$, and $M_{\mathrm{w}} = 10$, similar to the one chosen for the posterior verification of DPD and IBFD.

Table 3.2 Computational complexity required to execute the model calculation and coefficient update stages of the SPW identification algorithm, in terms of real multiplications per sample. Also a numerical example is presented, where $P = 3$, and $M_{\mathrm{w}} = 10$.

Operation		Symbolic real mul./sample	Numeric real mul./sample
Model calculation	$s[n]$	$4M_{\mathrm{w}}$	40
	$y[n]$	$P^2+4P+8+sqrt$	29
	Total	$P^2+4P+4M_{\mathrm{w}}+8+sqrt$	69
Coefficient update	$\mathbf{q}_{n+1}$	$2P+8$	14
	$\mathbf{w}_{n+1}$	$P^2+3P+6M_{\mathrm{w}}+15+div$	93
	Total	$P^2+5P+6M_{\mathrm{w}}+23+div$	107
Total per iteration		$2P^2+9P+10M_{\mathrm{w}}+31+div+sqrt$	176

3.4 Spline-based Wiener-Hammerstein Model

The processing steps and learning equations of the spline-based Wiener-Hammerstein model are next detailed, as detailed in [P3]. This model is a combination of the Wiener and Hammerstein approaches that further enhances the modeling capabilities by adding an extra memory filter, compensating for extra memory appearing

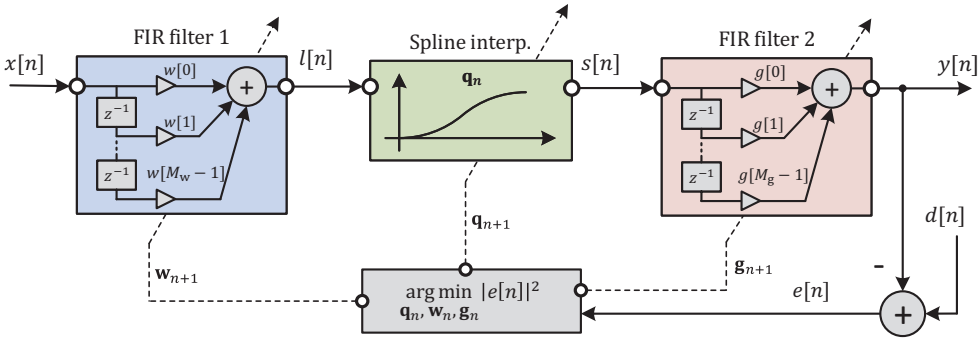


Figure 3.6 Complete architecture of the complex cascaded SPWH structure. The *spline interpolation block* comprises the processing scheme presented in Fig. 3.3.

in the system. The serial structure of this model, depicted in Fig. 3.6, contains first a linear filter, followed by a nonlinearity, and is further cascaded with another linear filter [33], [226]. Since the elementary nonlinear blocks are implemented with the presented spline interpolation scheme, this model is referred to as SPWH. In DPD applications, the linearization performance is further enhanced by the SPWH approach, as it better models the pre- and post-memory of the PA. In IBFD applications, the SPWH model is able to model the memory appearing before the TX PA, and also that of the SI multipath channel. These improvements, both in DPD and IBFD, come also with the cost of a slightly increased computational complexity.

Let us denote the input and output model signals as $x[n]$ and $y[n]$, the first intermediate signal as $l[n]$, and the second intermediate signal as $s[n]$, according to the nomenclature presented in Fig. 3.6. The intermediate and output signals, after applying a convolution for the FIR filter and the presented spline interpolated scheme for the nonlinear block, can be written as

$$l[n] = \mathbf{w}_n^T \mathbf{x}_n, \tag{3.30}$$

$$s[n] = l[n]\boldsymbol{\Psi}_n^T(\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_n), \tag{3.31}$$

$$y[n] = \mathbf{g}_n^T \mathbf{s}_n, \tag{3.32}$$

where $\mathbf{s}_n \in \mathbb{C}^{M_g \times 1} = [s[n]\ s[n-1]\ \cdots\ s[n-M_g+1]]^T$ is the signal regression of $s[n]$.

3.4.1 Learning Equations

In order to make the SPWH model adaptive, there are three parameters that need to be estimated, namely the spline control points, $\mathbf{q}_n$, and the two LTI FIR filters, $\mathbf{w}_n$ and $\mathbf{g}_n$. The error signal can be defined in the same way as in (3.18), but now the cost function depends on these three parameters to adapt, and can be expressed as

$$J(\mathbf{q}_n, \mathbf{w}_n, \mathbf{g}_n) = e[n]e^*[n], \tag{3.33}$$

where $e[n]$ remains as in (3.18). Taking the gradient-descent approach to minimize the cost function, the following learning rules are obtained, for each parameter to be learned,

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_{\mathrm{q}} \nabla_{\mathbf{q}_n} J(\mathbf{q}_n, \mathbf{w}_n, \mathbf{g}_n), \tag{3.34}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{\mathrm{w}} \nabla_{\mathbf{w}_n} J(\mathbf{q}_n, \mathbf{w}_n, \mathbf{g}_n), \tag{3.35}$$

$$\mathbf{g}_{n+1} = \mathbf{g}_n - \mu_{\mathrm{g}} \nabla_{\mathbf{g}_n} J(\mathbf{q}_n, \mathbf{w}_n, \mathbf{g}_n). \tag{3.36}$$

After evaluating the complex gradient operator, the final learning equations can be presented as

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_{\mathrm{q}} e[n] \boldsymbol{\Sigma}_n^T \mathbf{L}_n^* \mathbf{g}_n^*, \tag{3.37}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_{\mathrm{w}} \Big(e[n] \boldsymbol{\Gamma}_n \mathbf{g}_n^* + \boldsymbol{\Upsilon}_n \operatorname{Re}\Big\{ e[n] \dot{\mathbf{Y}}_n \mathbf{g}_n^* \Big\} \Big), \tag{3.38}$$

$$\mathbf{g}_{n+1} = \mathbf{g}_n + \mu_{\mathrm{g}} e[n] \mathbf{s}_n^*, \tag{3.39}$$

where $\mathbf{L}_n \in \mathbb{C}^{M_{\mathrm{g}} \times M_{\mathrm{g}}}$ and $\dot{\mathbf{Y}}_n \in \mathbb{C}^{M_{\mathrm{g}} \times M_{\mathrm{g}}}$ contain the signal regression of $l[n]$ and $\dot{y}[n]$ in their main diagonals, respectively, and $\boldsymbol{\Gamma}_n$ and $\boldsymbol{\Upsilon}_n$ are defined as

$$\boldsymbol{\Gamma}_n = \Big[\boldsymbol{\Psi}_n (\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_n^*) \mathbf{x}_n^* \; \boldsymbol{\Psi}_{n-1} (\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_{n-1}^*) \mathbf{x}_{n-1}^* \; \cdots \; \cdots \; \boldsymbol{\Psi}_{n-M_{\mathrm{g}}+1} (\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_{n-M_{\mathrm{g}}+1}^*) \mathbf{x}_{n-M_{\mathrm{g}}+1}^* \Big], \tag{3.40}$$

$$\boldsymbol{\Upsilon}_n = \begin{bmatrix} \frac{l[n]\mathbf{x}_n^*}{\Delta_l \lvert l[n] \rvert} & \frac{l[n-1]\mathbf{x}_{n-1}^*}{\Delta_l \lvert l[n-1] \rvert} & \cdots & \frac{l[n-M_{\mathrm{g}}+1]\mathbf{x}_{n-M_{\mathrm{g}}+1}^*}{\Delta_l \lvert l[n-M_{\mathrm{g}}+1] \rvert} \end{bmatrix}. \tag{3.41}$$

Note that the final forms of $\mathbf{q}_{n+1}$ and $\mathbf{g}_{n+1}$ correspond to the expressions obtained in the SPH case. In fact, the second filter, $\mathbf{g}_n$, can be removed (i.e., $\mathbf{g}_n = [1 \; 0 \; \cdots \; 0]^T$)

Table 3.3 Computational complexity required to execute the model calculation and coefficient update stages of the SPWH identification algorithm, in terms of real multiplications per sample. Also a numerical example is presented, where $P = 3$, $M_{\mathrm{w}} = 7$, $M_{\mathrm{g}} = 7$, and $\tau^{\mathrm{SP}} = 5$.

Operation		Symbolic real mul./sample	Numeric real mul./sample
Model calculation	$l[n]$	$4M_{\mathrm{w}}$	28
	$s[n]$	$P^2 + 4P + 8 + sqrt$	29
	$y[n]$	$4M_{\mathrm{g}}$	28
	Total	$P^2 + 4P + 4(M_{\mathrm{w}} + M_{\mathrm{g}}) + 8 + sqrt$	85
Coefficient update	$\mathbf{q}_{n+1}$	$4P + 2P\tau^{\mathrm{SP}} + 6\tau^{\mathrm{SP}} + 6$	78
	$\mathbf{w}_{n+1}$	$P^2 + 3P + 6M_{\mathrm{w}}M_{\mathrm{g}} + 14M_{\mathrm{w}} + 8M_{\mathrm{g}} + 9 + div$	475
	$\mathbf{g}_{n+1}$	$4M_{\mathrm{g}} + 2$	30
	Total	$P^2 + 7P + 2P\tau^{\mathrm{SP}} + 6\tau^{\mathrm{SP}} + 6M_{\mathrm{w}}M_{\mathrm{g}}$ $+14M_{\mathrm{w}} + 12M_{\mathrm{g}} + 17 + div$	583
Total per iteration		$2P^2 + 11P + 2P\tau^{\mathrm{SP}} + 6\tau^{\mathrm{SP}} + 6M_{\mathrm{w}}M_{\mathrm{g}}$ $+18M_{\mathrm{w}} + 16M_{\mathrm{g}} + 25 + sqrt + div$	668

in (3.38) to arrive at (3.29), and the nonlinear response $\mathbf{q}_n$ can be further removed (i.e., $\mathbf{q}_n = [0\,0\,\cdots\,0]^T$) to arrive at (3.23), the classical least mean squares (LMS) filter expression, as reviewed in [P3].

3.4.2 Complexity Analysis

The computational complexity of the SPWH model is analyzed in the same manner as done for the SPH and SPW approaches. It is divided into the model identification and the coefficient update stages, and expressed in terms of real multiplications per sample. Since the SPWH model shares the same update for the control point coefficients as the SPH model, the same complexity-reduction method is utilized, and hence the complexity expressions are left as a function of the reduced time span τ^{SP}.

The processing complexity of the SPWH model follows the processing steps presented in the previous subsections, and is presented in Table 3.3, as a function of the SPWH model parameters. A numerical example is also included in the table, when selecting $P = 3$, $M_{\mathrm{w}} = 7$, $M_{\mathrm{g}} = 7$, and $\tau^{\mathrm{SP}} = 5$. The SPWH model benefits from an enhanced modeling performance, but the required computational complexity is also increased, as one additional cascaded block and its corresponding learning equation are newly added.

3.5 Spline-based Hammerstein-Wiener Model

This section presents the last cascaded model of the chapter, which is the spline-based Hammerstein-Wiener model, named SPHW herein. The SPHW model is a combination of the Hammerstein and Wiener models, thus its structure, depicted in Fig. 3.7, contains first an elementary nonlinear block, followed by a linear filter, and is ended with a second nonlinear block [57], [226], [242]. As in previous sections, the nonlinear blocks are implemented with the spline-interpolated scheme, and the linear block is built with an FIR filter. The modeling capabilities of the SPHW model are generally better than the SPH and SPW models alone, but also at the cost of an increased computational complexity. As demonstrated in [P7], this structure is particularly designed to modeling IBFD devices, since the first nonlinear block accounts for the effects of the TX PA, the linear filter models the SI multipath channel, and the second nonlinear block further compensates for the nonlinear effects of the RX LNA [12], [147]. Therefore, in the validation chapter, this cascaded model is only tested in the context of IBFD applications.

With reference to Fig. 3.7, and denoting $x[n]$ as the input model signal, the expressions of the output model signal, $y[n]$, and the intermediate signals, $l[n]$ and $s[n]$, can be expressed according to the derived spline interpolation scheme and the FIR convolution as

$$l[n] = x[n]\mathbf{\Psi}_n^T(\mathbf{1}_{Q^{\text{SP}}} + \mathbf{q}_n), \tag{3.42}$$

$$s[n] = \mathbf{w}_n^T \mathbf{l}_n, \tag{3.43}$$

$$y[n] = s[n]\mathbf{\Phi}_n^T(\mathbf{1}_{C^{\text{SP}}} + \mathbf{c}_n), \tag{3.44}$$

where $\mathbf{l}_n \in \mathbb{C}^{M_\text{w}\times 1} = [l[n]\ l[n-1]\ \cdots\ l[n-M_\text{w}+1]]^T$ is the signal regression of $l[n]$, $\mathbf{c}_n \in \mathbb{C}^{C^{\text{SP}}\times 1} = [c_0\ c_1\ \cdots\ c_{C^{\text{SP}}-1}]^T$ contains the control points for the second spline-interpolated scheme, and $\mathbf{\Phi}_n$ is equivalent to (3.6), built by mimicking the expressions presented in (3.10), (3.11), and (3.7), and considering a spline order K.

3.5.1 Learning Equations

The learning equations of the SPHW model can also be extracted by using the gradient-descent approach, applied to the cost function. The cost function now depends

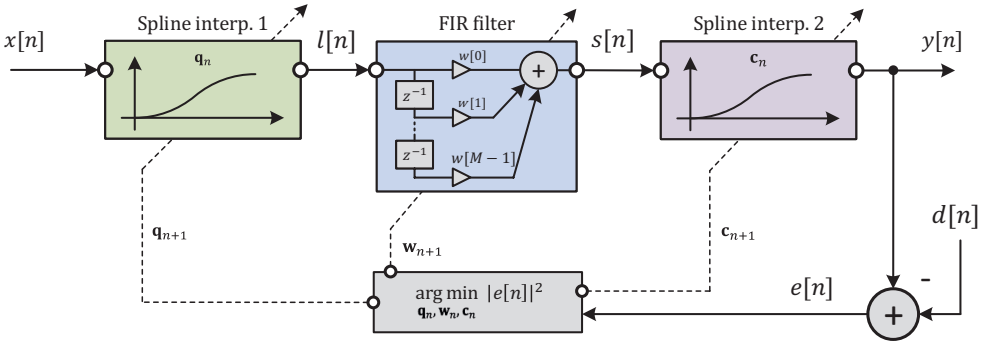


Figure 3.7 Complete architecture of the complex cascaded SPHW structure. The *spline interpolation block* comprises the processing scheme presented in Fig. 3.3.

on the parameters $\mathbf{q}_n$, $\mathbf{c}_n$, and $\mathbf{w}_n$, and its defined as

$$J(\mathbf{q}_n, \mathbf{c}_n, \mathbf{w}_n) = e[n]e^*[n], \tag{3.45}$$

where $e[n]$ remains as in (3.18). Applying the gradient-descent solution, the learning rules can be initially formulated as

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_{\mathrm{q}} \nabla_{\mathbf{q}_n} J(\mathbf{q}_n, \mathbf{c}_n, \mathbf{w}_n), \tag{3.46}$$

$$\mathbf{c}_{n+1} = \mathbf{c}_n - \mu_{\mathrm{c}} \nabla_{\mathbf{c}_n} J(\mathbf{q}_n, \mathbf{c}_n, \mathbf{w}_n), \tag{3.47}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{\mathrm{w}} \nabla_{\mathbf{w}_n} J(\mathbf{q}_n, \mathbf{c}_n, \mathbf{w}_n), \tag{3.48}$$

where μ_{c} is the learning rate for the second nonlinearity. Developing these expressions, the final learning equations for the SPHW model are obtained, reading

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_{\mathrm{q}} e[n] \boldsymbol{\Sigma}_n^T \mathbf{X}_n^* \mathbf{w}_n^* \boldsymbol{\Phi}_n^T (\mathbf{1}_{C^{\mathrm{SP}}} + \mathbf{c}_n^*), \tag{3.49}$$

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu_{\mathrm{c}} e[n] s^*[n] \boldsymbol{\Phi}_n, \tag{3.50}$$

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_{\mathrm{w}} \mathbf{l}_n^* \left(e[n] \boldsymbol{\Phi}_n^T (\mathbf{1}_{Q^{\mathrm{SP}}} + \mathbf{q}_n^*) + \frac{2s[n] \mathrm{Re}\{e[n]\dot{y}[n]\}}{\Delta_s |s[n]|} \right). \tag{3.51}$$

It is finally noted that these learning rules can be updated independently, depending on the final application in which the SPHW algorithm is executed.

Table 3.4 Computational complexity required to execute the model calculation and coefficient update stages of the SPHW identification algorithm, in terms of real multiplications per sample. Also a numerical example is presented, where $P = 3$, $K = 3$, $M_{\mathrm{w}} = 10$, and $\tau^{\mathrm{SP}} = 5$.

Operation		Symbolic real mul./sample	Numeric real mul./sample
Model calculation	$l[n]$	$P^2 + 4P + 8 + sqrt$	29
	$s[n]$	$4M_{\mathrm{w}}$	40
	$y[n]$	$K^2 + 4K + 8 + sqrt$	29
	Total	$P^2 + K^2 + 4P + 4K + 4M_{\mathrm{w}} + 16 + 2sqrt$	98
Coefficient update	$\mathbf{q}_{n+1}$	$2K + 8$	14
	$\mathbf{c}_{n+1}$	$4P + 2P\tau^{\mathrm{SP}} + 6\tau^{\mathrm{SP}} + 10$	82
	$\mathbf{w}_{n+1}$	$K^2 + 3K + 6M_{\mathrm{w}} + 15 + div$	93
	Total	$4P + 2P\tau^{\mathrm{SP}} + 6\tau^{\mathrm{SP}} + K^2$ $+5K + 6M_{\mathrm{w}} + 33 + div$	189
Total per iteration		$P^2 + 8P + 2P\tau^{\mathrm{SP}} + 6\tau^{\mathrm{SP}} + 2K^2$ $+9K + 10M_{\mathrm{w}} + 49 + div + 2sqrt$	287

3.5.2 Complexity Analysis

The computational complexity of the SPHW model is also divided into the stages of model identification and coefficient estimation, and it is expressed in terms of real multiplications per sample. The SPHW model shares the same update as the SPH model for the first spline-interpolated scheme (i.e., $\mathbf{q}_n$), thus the same complexity reduction method is considered, leaving the final complexity expressions as a function of the reduced time span τ^{SP}.

The final processing complexity expressions corresponding the SPHW model are then presented in Table 3.4, as a function of the SPHW mode parameters. The table also includes a numerical example, when selecting $P = 3$, $K = 3$, $M_{\mathrm{w}} = 10$, and $\tau^{\mathrm{SP}} = 5$, a similar parametrization to the one chosen in for the experimental verification of the models.

4 LINEAR-IN-PARAMETERS MODELS

THIS chapter presents a collection of linear-in-parameters behavioral models to be used in digital predistortion (DPD) and digital SI cancellation (DSIC) applications, alongside with two alternative parameter learning solutions. Additionally, different complexity-reduction techniques are presented, which aim at easing the parameter learning procedure. These techniques include several sign-based solutions applied to classical closed-loop learning methods, and also various low-complexity inverse covariance matrix (ICM) estimation solutions.

The contents of this chapter can be found in the publications [P1, P2, P4, P6, P8], as well as in the work presented in [200].

4.1 Spline-based MP-LUT Model

This section overviews the developed spline-based MP-LUT model, which is essentially a combination between an MP model, implemented with LUTs, and the spline interpolation scheme reviewed in Section 3.1. Inspired by the works proposed and discussed in [180], [189], an MP parallel-branched structure is designed to model the memory effects, while the nonlinearities in each parallel branch are substituted with the complex spline interpolation scheme presented above. For the remainder of the thesis, this model is referred to as SPMP, and its complete architecture is conceptually depicted in Fig. 4.1. More elaborated details of the SPMP model can be found in [P1]. The novelty and contributions of this section can be found in the combination of the MP-LUT processing architecture with the injection-based spline interpolation scheme, together with the derived learning equations which are adapted following the complex gradient-descent algorithm. A complexity analysis of the resulting SPMP model can also be found at the end of this section, which details both model identification and coefficient learning complexities.

In terms of modeling capabilities, SPMP is generally a richer model when com-

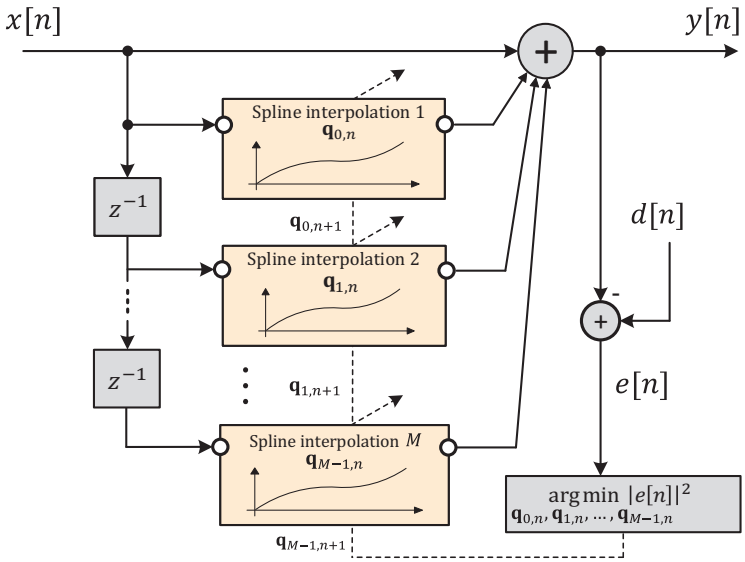


Figure 4.1 Complete architecture of the complex-valued memory polynomial-type SPMP structure. Each of the branches incorporates the spline interpolation scheme presented in Section 3.1. Additionally, the *spline interpolation block* comprises the processing scheme presented in Fig. 3.3.

pared to the cascaded approaches reviewed in Chapter 3, since an individual nonlinear function is given for each memory branch. This method, on the other hand, also entails higher complexity, especially when the considered number of memory taps is large. When compared to canonical MP models, the SPMP approach poses several advantages. First of all, the model identification processing complexity is greatly reduced when compared to classical MP solutions, since the parallel LUTs provide a very simple approach to process the otherwise higher order polynomial functions [P1]. Secondly, the learning path computational complexity is particularly reduced, owing to the utilization of gradient-descent techniques in combination with the interpolated LUTs. These derived learning approaches avoid the need for using extra BFs pre-whitening or orthogonalization [4], [200], or other self-orthogonalized learning procedures [189], [251], which drastically simplifies the involved learning complexities. Canonical MP models indeed require BF orthogonalization or pre-whitening due to the poor convergence performance of the LMS algorithm if the elements of the input vector are highly correlated, stemming from the large eigenvalue spread of the input signal CM. In any such model, the static BFs

are certainly correlated, as they are nonlinear functions of the transmitted signal. A third advantage of the SPMP model – which constitutes also a piece-wise approach due to the LUT interpolation scheme – is the ability to use different interpolation orders for different regions, depending on the desired modeling accuracy in each.

4.1.1 SPMP Processing Scheme

Following the nomenclature chosen in Fig. 4.1, and the complex injection-based spline interpolation scheme presented in Section 3.1, the input-output relation of the SPMP model can be directly expressed as [P1]

$$y[n] = x[n] + \sum_{m=0}^{M-1} x[n-m]\boldsymbol{\Psi}_{n-m}^{T}\mathbf{q}_m, \tag{4.1}$$

where $x[n]$ and $y[n]$ are the input and output model signals, M denotes the number of memory taps, $\boldsymbol{\Psi}_n$ is defined in (3.6), with also a spline interpolation order P, and $\mathbf{q}_m$, $m = 0,\ 1,\ \cdots,\ M-1$, are the M LUTs of the model, one for each branch. Each of the LUTs contains the corresponding control points used in the spline-interpolated scheme to model each nonlinearity. It is finally noted that a different spline parametrization could be adopted in each branch, for instance choosing a more powerful interpolation scheme for the instantaneous or the nearest past memory branches, which are typically the ones contributing more to the model output. For simplicity, however, the examples provided in the experimental validation chapter consider the same spline order for every branch. A more detailed development of the SPMP system model can be found in [P1].

4.1.2 Learning Equations

The complex-valued learning equations of the SPMP model are next derived. In contrast to the cascaded models presented in Chapter 3, the SPMP model does not contain any type of cascaded filters, and hence the learning entity considers the individual control points appearing in the M parallel spline-interpolated branches, $\mathbf{q}_{m,n}$, $m = 0,\ 1,\ \cdots,\ M-1$. Since in this work the spline interpolation scheme is the same in every branch, a single learning equation is obtained, and then applied M times to update each control point vector in each branch.

With the goal of deriving the corresponding learning expressions, an instantaneous error signal can be first defined as $e[n] = d[n] - y[n]$, and the corresponding cost function for the SPMP model can be written as

$$J(\mathbf{q}_{0,n},\, \mathbf{q}_{1,n},\, \cdots,\, \mathbf{q}_{M-1,n}) = e[n]e^{*}[n]. \tag{4.2}$$

Applying the complex-valued steepest gradient-descent solution to the mth control point vector of the cost function, its learning expression can be defined as

$$\mathbf{q}_{m,n+1} = \mathbf{q}_{m,n} - \mu_{\mathrm{q}_m} \nabla_{\mathbf{q}_{m,n}} J(\mathbf{q}_{0,n}, \mathbf{q}_{1,n}, \cdots, \mathbf{q}_{M-1,n}), \tag{4.3}$$

where μ_{q_m} is the corresponding learning step for each branch. Further developing this expression leads to the final learning rule, which reads

$$\mathbf{q}_{m,n+1} = \mathbf{q}_{m,n} + \mu_{\mathrm{q}_m} e[n] x^{*}[n-m]\, \mathbf{\Psi}_{n-m}. \tag{4.4}$$

This learning equation is then executed to update the M parallel branches of the SPMP model.

4.1.3 Complexity Analysis

The computational complexity of the SPMP model is presented next. As in Chapter 3, it is divided into the stages of model identification and coefficient estimation, and is presented in terms of real multiplications per sample. In the model identification stage, the complexity is composed by the sum of the M parallel spline-interpolated branches, with input samples $x[n-m]$, $m = 0,\, 1,\, \cdots,\, M-1$, as shown in Fig. 4.1. It is additionally noted that, at time instant n, only the instantaneous spline weight $\mathbf{\Psi}_n$ needs to be calculated, as the past vectors $\mathbf{\Psi}_{n-1}\ \cdots \mathbf{\Psi}_{n-M+1}$ can be obtained from previous sample instants. The complexity in the learning stage comes from the derived learning equation in (4.4), which is executed for each memory branch, thus exactly M times.

The final overall complexity expressions of the SPMP model are then presented in Table 4.1, as a function of the SPMP model parameters. These expressions are calculated following the exact processing steps presented in Section 4.1.

Table 4.1 Computational complexity required to execute the model identification and coefficient update stages of the SPMP identification algorithm, in terms of real multiplications per sample.

Operation		Symbolic real mul./sample	Numeric real mul./sample
Model identification	$y[n]$	$P^2+3P+2PM+6M+4$	142
	Total	$P^2+3P+2PM+6M+4$	142
Coefficient update	$\mathbf{q}_{m,n+1}$	$M(2P+8)$	140
	Total	$M(2P+8)$	140
Total per iteration		$P^2+3P+4PM+14M+4$	282

4.2 MP-LUT Model

The MP-LUT model is presented next. This approach is a special case of the previously presented SPMP model, and also the work presented in [189]. Specifically, this model is configured with a nearest neighbor interpolation scheme (i.e., $P=0$ in the spline processing), and it is further presented in a matrix notation equivalent form, that will ease its manipulation in further DPD experiments. The motivation behind removing the spline interpolation scheme comes from its inherent reduced complexity [180], [207], while the solution still achieves similar levels of performance, as demonstrated in the next chapter, albeit that a larger number of LUT entries is required. Also, the non-interpolated scheme is an important ingredient if the SRA or sign-sign algorithms are to be used in combination with the learning rates, as the non-interpolated LUTs will not result in rank-deficiency problems when signing the input BF matrix. This concept is further explored in Section 4.3 and publications [P2,P6].

The usefulness of this model can be specifically found in the associated modified closed-loop learning algorithms of the MP-LUT model. First, classical GN, SO, and BLMS learning algorithms are shown in Section 4.2.2 to provide an overview of the already studied state-of-the-art techniques. Second, several sign-based solutions (Section 4.3) are applied to the GN, SO, and BLMS algorithms in order to reduce their computational complexity. In the next chapter, the performance-complexity trade-offs of each resulting sign-based algorithm are also reported to show how they could fit in different applications. Third, several ICM estimation methods (Section 4.4) are derived and applied to the SO learning rule. The ICM is an important ingredient for the proper operation of the SO rule, and it needs to be recursively calculated. These

methods are able to reduce the computational complexity of calculating this matrix, which is typically the heaviest term in the learning update. Overall, the proposed methods allow for remarkable complexity reductions in the learning update of the MP-LUT model.

4.2.1 MP-LUT Processing Scheme

The conceptual architecture of the MP-LUT model is equivalent to the one presented for the SPMP model in Fig 4.1, but instead considers a 0th-degree spline interpolation (i.e., $P = 0$). Thereby, its input-output relation, following the same notation as in the Fig 4.1, can be also written as

$$y[n] = x[n] + \sum_{m=0}^{M-1} x[n-m]\boldsymbol{\Psi}_{n-m}^{T}\mathbf{q}_m, \tag{4.5}$$

where $\boldsymbol{\Psi}_n \in \mathbb{R}^{Q\times 1} = [0 \cdots 0\ 1\ 0 \cdots 0]^T$ now, since $P = 0$ is chosen for the spline interpolation scheme. The nonzero element in $\boldsymbol{\Psi}_n$ is still indexed in the i_nth position, according to (3.10), and $\mathbf{q}_m$, $m = 0,\ 1,\ \cdots M-1$ are the Q-sized non-interpolated M LUTs of the model.

For easiness in posterior DPD manipulation, the form in (4.5) can alternatively be expressed following matrix notation. Such notation contains the complete input BF matrix and a column vector stacking the M LUTs of the model. This form can be written as

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\Omega}\boldsymbol{\alpha}, \tag{4.6}$$

where $\mathbf{x} \in \mathbb{C}^{N\times 1} = [x[n]\ x[n-1] \cdots x[n-N+1]]^T$ and $\mathbf{y} \in \mathbb{C}^{N\times 1} = [y[n]\ y[n-1] \cdots y[n-N+1]]^T$ denote the input and output data vectors, $\boldsymbol{\Omega} \in \mathbb{C}^{N\times C}$ is the input BF matrix, whose form is presented in (4.7), $C = MQ^{\mathrm{SP}}$, and $\boldsymbol{\alpha} \in \mathbb{C}^{C\times 1} = [\alpha_0\ \alpha_1 \cdots \alpha_{C-1}]^T$ contains the M LUTs of each polynomial branch. Note that the formulation in (4.6) also follows the injection-based principle, and the input BF matrix presented in (4.7) can be modified accordingly to also fit other Volterra-based models, such as the well-known GMP method.

$$\boldsymbol{\Omega} = \begin{bmatrix} x[n]\boldsymbol{\Psi}_n^T & x[n-1]\boldsymbol{\Psi}_{n-1}^T & \cdots & x[n-M]\boldsymbol{\Psi}_{n-M}^T \\ x[n+1]\boldsymbol{\Psi}_{n+1}^T & x[n]\boldsymbol{\Psi}_n^T & \cdots & x[n-M+1]\boldsymbol{\Psi}_{n-M+1}^T \\ \vdots & \vdots & \ddots & \vdots \\ x[n+N-1]\boldsymbol{\Psi}_{n+N-1}^T & x[n+N-2]\boldsymbol{\Psi}_{n+N-2}^T & \cdots & x[n+N-M-1]\boldsymbol{\Psi}_{n+N-M-1}^T \end{bmatrix} \tag{4.7}$$

4.2.2 Classical Closed-loop Learning Algorithms

In this subsection, classical closed-loop learning algorithms are reviewed. Those algorithms are already presented in current literature, and serve as a basis to apply the upcoming reduced-complexity signed algorithms and ICM estimation methods.

By formulating the proposed MP-LUT model as a linear-in-parameters model, with the processing structure presented in (4.6), closed-loop learning algorithms can be applied. Specifically, the damped GN [108], [109], [189], [258], SO [4], [200], [251], and BLMS [84], [236], [259] closed-loop learning techniques are discussed and analyzed within this work. First, the block-based error signal, $\mathbf{e}_k$, is defined, at iteration k and for a K^{it}-sized block of samples, as

$$\mathbf{e}_k = \mathbf{x}_k - \mathbf{y}_k, \tag{4.8}$$

where $\mathbf{x}_k$ and $\mathbf{y}_k$ are now the K^{it}-sized input and output data blocks. The learning equations can be consequently defined as

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{GN}}\left(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k\right)^{-1} \boldsymbol{\Omega}_k^H \mathbf{e}_k, \tag{4.9}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{SO}}(\mathbf{R}^*)^{-1} \boldsymbol{\Omega}_k^H \mathbf{e}_k, \tag{4.10}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{BL}} \boldsymbol{\Omega}_k^H \mathbf{e}_k, \tag{4.11}$$

where μ_{GN}, μ_{SO}, and μ_{BL} are the learning rates for the GN, SO, and BLMS learning rules, $\boldsymbol{\Omega}_k \in \mathbb{C}^{K^{\text{it}} \times C}$, and K^{it} is the iteration estimation sample size. Additionally, $\mathbf{R} \in \mathbb{C}^{C \times C} = \text{E}\{\boldsymbol{\upsilon}[n]\boldsymbol{\upsilon}^H[n]\}$ is the input data covariance matrix [263], considering $\boldsymbol{\upsilon}[n] \in \mathbb{C}^{C \times 1}$ is the instantaneous input data vector, which can be seen as the first row of the BF matrix $\boldsymbol{\Omega}$. The CM can be estimated by using the classical sample

estimation method as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{\upsilon}[n]\boldsymbol{\upsilon}^{H}[n]. \tag{4.12}$$

It is worth noting the differences found in the GN, SO, and BLMS learning equations. The first equation, corresponding to GN, executes the term $\left(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k\right)^{-1}$ in each learning iteration. This results in an enhanced performance, since the statistics of the input signal are well-characterized at each iteration, but its continuous calculation also implies very high complexity numbers. Secondly, the SO solution estimates the input data CM only once, and this estimate then remains fixed for as long as the input signal remains unmodified, within specific limits [251]. How often to update the input covariance matrix (CM) needs to be decided according to the performance and complexity goals of the final application hosting the algorithm. In general, this estimation can remain fixed for a large number of iteration, as demonstrated in [P4, P8] , and the complexity is usually reduced when compared to GN. Finally, the BLMS learning equation can be considered to use an identity CM, constituting already a very crude approximation for the term $\left(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k\right)^{-1}$. Obviously, this leads to drastically reduced complexity numbers, but implies a reduced modeling performance, and also a slow convergence rate. In many cases, the convergence speed of the BLMS model is so slow that it is unfeasible to be used in commercial applications, as was reviewed in [P2, P6] . Orthogonalization or pre-whitening methods can be applied to increase the convergence speed of the model, but these methods, on the other hand, also increase the complexity of the learning equation.

4.3 Closed-loop Signed Learning Algorithms

With the aim of reducing the computational complexity of the aforementioned learning algorithms in (4.9) (GN), (4.10) (SO), and (4.11) (BLMS), computationally efficient sign-based versions of the previous rules are formulated. These methods are presented in detail in [P2, P6] , and sign different terms in the original learning rules such that the final magnitudes of the real and imaginary parts are reduced to either +1 or -1. Operations with these magnitudes are trivial to execute in DSP devices [237], thus the proposed algorithms are able to greatly reduce the overall number of required multiplications.

Considering complex-valued I/Q data, the classical definition of the sign function projects a nonzero number into the unit circle of the complex plane; this is $\{z \in \mathbb{C} \mid |z| = 1\}$, as reviewed in [239]. The magnitude of the resulting complex number, $|z|$, is equal to $+1$, but the individual real and imaginary parts are not necessarily either $+1$ or -1, thus no complexity reduction is yet achieved when multiplying with z. In order to remove the need for multiplications, the sign function is instead defined as

$$\operatorname{csgn}(z) := \operatorname{sgn}(\operatorname{Re}(z)) + j \operatorname{sgn}(\operatorname{Im}(z)), \tag{4.13}$$

which results in either $+1$ or -1 for both real and imaginary parts, hence avoiding the need for computing multiplications when multiplying with z. For matrices, this complex-valued signum function is done element-wise, such that all elements are signed. The following sections present the proposed signed-based learning rules, which are specifically combined with the sign, sign regressor, and sign-sign algorithms.

4.3.1 Sign Algorithm

The sign algorithm is the first signed-based method reviewed herein. It can be directly obtained by signing the block error signal in the learning rules presented in (4.9), (4.10), and (4.11) [246]. With such an approach, multiplications in the term $\mathbf{\Omega}_k^H \mathbf{e}_k$ are avoided in the GN, SO, and BLMS rules. This is already quite a substantial reduction, since the horizontal dimension (i.e., number of columns) of $\mathbf{\Omega}_k^H$ is typically large in DPD implementations. To provide a numerical example, in the experiments testing the sign-based algorithms, K^{it} is already $25,000$ samples.

By signing the error vector, $\mathbf{e}_k$, the sign learning equations read

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{GN}} \left(\mathbf{\Omega}_k^H \mathbf{\Omega}_k\right)^{-1} \mathbf{\Omega}_k^H \operatorname{csgn}(\mathbf{e}_k), \tag{4.14}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{SO}} (\mathbf{R}^*)^{-1} \mathbf{\Omega}_k^H \operatorname{csgn}(\mathbf{e}_k), \tag{4.15}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{BL}} \mathbf{\Omega}_k^H \operatorname{csgn}(\mathbf{e}_k). \tag{4.16}$$

The complexity of this method is analyzed later in this section.

4.3.2 Sign Regressor Algorithm

The SRA can be obtained by signing the Hermitian transpose BF matrix, $\mathbf{\Omega}_k^H$, in the GN and BLMS learning rules. Thereby, multiplications in the terms $\mathbf{\Omega}_k^H\mathbf{\Omega}_k$ and $\mathbf{\Omega}_k^H\mathbf{e}_k$ (GN) and in the term $\mathbf{\Omega}_k^H\mathbf{e}_k$ (BLMS) are avoided, reducing the overall computational complexity of the learning solutions. In the GN method, the complexity savings are increased when compared to the sign algorithm, as one extra term is signed. The complexity of the BLMS method, in turn, remains the same, thus the best working solution in terms of performance should be selected.

The resulting SRA versions of the GN and BLMS read

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{GN}}\Big(\operatorname{csgn}(\mathbf{\Omega}_k^H)\mathbf{\Phi}_k\Big)^{-1}\operatorname{csgn}(\mathbf{\Omega}_k^H)\mathbf{e}_k, \tag{4.17}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{BL}}\operatorname{csgn}(\mathbf{\Omega}_k^H)\mathbf{e}_k. \tag{4.18}$$

When considering the SO learning solution, the SRA method cannot be applied as such, since the corresponding learning rule already contains the ICM, which is the estimation of the term $\mathbf{\Omega}_k^H\mathbf{\Omega}_k$, where the SRA would have been originally applied. Two alternative solutions for combining the SRA with the SO rule are proposed instead. The first form only signs the data matrix $\mathbf{\Omega}_k^H$, hence avoiding the calculation of the term $\mathbf{\Omega}_k^H\mathbf{\Omega}_k$. The second form signs the ICM as a whole, avoiding the calculation of the term $(\mathbf{R}^*)^{-1}\mathbf{\Omega}_k^H$. Both of these approaches read

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{SO}}(\mathbf{R}^*)^{-1}\operatorname{csgn}(\mathbf{\Omega}_k^H)\mathbf{e}_k, \tag{4.19}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{SO}}\operatorname{csgn}((\mathbf{R}^*)^{-1})\mathbf{\Omega}_k^H\mathbf{e}_k. \tag{4.20}$$

With the former SRA SO form, the reduced complexity is the same as in the sign SO case, as an equal number of multiplications is avoided. With the latter form, the achieved reduction is less than in the sign SO case. It is known from [P2] that the second form does not perform well in the experimental results carried out later, as signing the whole ICM is already a very crude approximation for this term. Considering this issue, and also that the complexity reduction is even less than in the sign SO case, this form is discarded in the experimental verification cases shown in Chapter 5.

At this point, it is important to note that all polynomial-type models, as well

as interpolated LUT approaches will suffer from a rank deficiency when applying the signum function to the term $\mathbf{\Omega}_k^H$ [109]. In polynomial-type models, repeated columns will appear in this matrix, and in interpolated LUT approaches, linear dependencies between columns will be generated. When this happens, the learning equations are not able to converge to any value, as they have multiple solutions. A specific example of this issue was presented in [109], where the authors needed to apply an extra Walsh-Hadamard transformation to gaussianize the input data BF matrix, $\mathbf{\Omega}_k^H$, in order to solve this problem. This matrix transformation, however, further increased the computational complexity of the learning equation. Nevertheless, the proposed non-interpolated MP-LUT model allows to directly apply the SRA algorithm as such, not producing any repeated columns in the input BF matrix $\mathbf{\Omega}_k^H$, as demonstrated in [P2]. With this approach, rank deficiency issues, and more importantly, extra matrix transformations that further increase the complexity of the learning equation, are avoided. This is another obvious benefit of the proposed MP-LUT model, in contrast to polynomial-type approaches.

4.3.3 Sign-sign Algorithm

The last sign-based algorithm studied within this work is the sign-sign algorithm. Such a method can be obtained by signing both the data matrix and the error term in the learning equations. Further complexity savings are obtained in the learning methods, which avoid a great number of multiplications. In the GN method, the complexity is further reduced, as only a few multiplications are required. In the BLMS method, the complexity, in terms of multiplications, is already zero. The sign-sign expressions for both GN and BLMS approaches read

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{GN}} \Big(\mathrm{csgn}(\mathbf{\Omega}_k^H)\mathbf{\Omega}_k \Big)^{-1} \mathrm{csgn}(\mathbf{\Omega}_k^H)\,\mathrm{csgn}(\mathbf{e}_k), \tag{4.21}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{BL}}\,\mathrm{csgn}(\mathbf{\Omega}_k^H)\,\mathrm{csgn}(\mathbf{e}_k). \tag{4.22}$$

The same conclusions can be drawn for the SO method, since the sign-sign algorithm cannot be applied to the SO rule as such. Again, two alternatives of this

method are proposed as

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{SO}} (\mathbf{R}^*)^{-1} \operatorname{csgn}(\boldsymbol{\Omega}_k^H) \operatorname{csgn}(\mathbf{e}_k), \tag{4.23}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\text{SO}} \operatorname{csgn}((\mathbf{R}^*)^{-1}) \boldsymbol{\Omega}_k^H \operatorname{csgn}(\mathbf{e}_k). \tag{4.24}$$

With both approaches, the final complexity is reduced to almost zero. However, similar conclusions as drawn before apply also for the latter SO form. Hence, it is presented herein for completeness of the study, but discarded in the posterior measurements carried out in Chapter 5.

It is finally noted that the same rank deficiency issue already appearing in the SRA case applies here as well, since the input data BF matrix $\boldsymbol{\Omega}_k^H$ needs to be signed in the same way. In other words, polynomial-type and interpolated LUT approaches will require extra matrix transformations, while the proposed MP-LUT model can be directly combined with the sign-sign algorithm as such.

4.3.4 Complexity Analysis

Table 4.2 Computational complexity analysis of the classical and signed-based GN, SO, and BLMS learning solutions, in terms of real multiplications and real additions per K^{it}-sized learning iteration.

Learning method	Symbolic real multiplications	Real additions
Gauss-Newton	$C^3 + 4M^2(K^{\text{it}}+1) + 2M(2K+1)$	$2M^2(K^{\text{it}}-1) + 2M(K^{\text{it}}+M-2) + 2C$
Sign Gauss-Newton	$C^3 + 4M^2(K^{\text{it}}+1) + 2C$	$2M^2(K^{\text{it}}-1) + 2M(K^{\text{it}}+M-2) + 2C$
SRA Gauss-Newton	$C^3 + 4M^2 + 2M$	$2M^2(K^{\text{it}}-1) + 2M(K^{\text{it}}+M-2) + 2C$
Sign-sign Gauss-Newton	$C^3 + 2M$	$2M^2(K^{\text{it}}-1) + 2M(K^{\text{it}}+M-2) + 2C$
Self-orthogonalization	$4(MN+M^2) + 2C$	$2M(K^{\text{it}}+C-2) + 2C$
Sign self-orthogonalization	$4CM + 2C$	$2M(K^{\text{it}}+C-2) + 2C$
SRA 1 self-orthogonalization	$4CM + 2C$	$2M(K^{\text{it}}+C-2) + 2C$
SRA 2 self-orthogonalization	$4MN + 2C$	$2M(K^{\text{it}}+C-2) + 2C$
Sign-sign 1 self-orthogonalization	$2C$	$2M(NM+1)$
Sign-sign 2 self-orthogonalization	$2C$	$2M(K^{\text{it}}+C-2) + 2C$
Block-LMS	$2M(2N+1)$	$2(MK^{\text{it}}+C)$
Sign block-LMS	$2M$	$2(MK^{\text{it}}+C)$
SRA block-LMS	$2M$	$2(MK^{\text{it}}+C)$
Sign-sign block-LMS	0	$2(MK^{\text{it}}+C)$

Table 4.3 Exact numerical complexity of the baseline and sign-based learning equations, when considering $M = 4$, $Q = 32$, $C = MQ^{\text{SP}} = 128$, and $K^{\text{it}} = 25,000$. The percentage reduction with respect each original learning equation is also shown in the table.

Learning method	Real multiplications	Multiplication reduction	Real additions
Gauss-Newton	4×10^6	0%	1×10^6
Sign Gauss-Newton	3.7×10^6	7.5%	1×10^6
SRA Gauss-Newton	2×10^6	50%	1×10^6
Sign-sign Gauss-Newton	2×10^6	50%	1×10^6
Self-orthogonalization	401×10^3	0%	201×10^3
Sign Self-orthogonalization	2.3×10^3	$> 99\%$	201×10^3
SRA 1 Self-orthogonalization	2.3×10^3	$> 99\%$	201×10^3
SRA 2 Self-orthogonalization	400×10^3	$< 1\%$	201×10^3
Sign-sign 1 Self-orthogonalization	256	$> 99\%$	800×10^3
Sign-sign 2 Self-orthogonalization	256	$> 99\%$	201×10^3
Block-LMS	400×10^3	0%	200×10^3
Sign Block-LMS	8	$> 99\%$	200×10^3
SRA Block-LMS	8	$> 99\%$	200×10^3
Sign-sign Block-LMS	0	100%	200×10^3

Once the sign-based learning equations are derived, a complexity analysis showing the expressions of the required multiplications and also the number of real additions per K^{it}-sized learning iteration is presented. The same assumptions as explained in previous sections regarding the calculation of complex multiplications are taken. It is also assumed that multiplications by ± 1 are free in DSP implementations [237]. Hence, following the exact processing steps of the baseline learning equations showed in (4.9), (4.10), and (4.11), and their corresponding sign-based versions, the final learning complexity expressions are presented in Table 4.2. Additionally, Table 4.3 presents the exact numerical complexity when the MP-LUT parameters take the example values of $M = 4$, $Q^{\text{SP}} = 32$, $C = MQ^{\text{SP}} = 128$, and $K^{\text{it}} = 25,000$. For convenience, these values are similar to the ones taken when performing the RF verification in Chapter 5. This table also presents the percentage reduction achieved with respect to the original baseline learning equations, in order to highlight the effectiveness of the proposed solutions. All in all, the sign-based algorithms are able to drastically reduce the required number of real multiplications in the learning stage compared to the original rules. Specifically, the achieved percentage reduction goes up to 50% in the case of GN, and more than 99% in the cases

of SO and BLMS, where the number of multiplications is even almost zero in some cases.

From Tables 4.2 and 4.3, four key conclusions can be extracted:

- The GN SRA and GN sign-sign methods mutually pose the same computational complexity, thereby the method providing the best performance should be selected.
- The SO sign and SO SRA 1 methods mutually pose the same computational complexity, thereby the method providing the best performance should be selected.
- The SO SRA 2 method does not provide any essential complexity reduction when compared to the original SO solution. It is also anticipated that the performance of this method lags behind the rest of SO solutions, thus it is discarded in the posterior RF verifications.
- The BLMS sign and BLMS SRA methods mutually pose the same computational complexity, thereby the method providing the best performance should be selected.

4.4 ICM Estimation Methods in SO Learning Algorithm

In this section, several methods which aim at efficiently estimating the ICM needed in the SO algorithms are presented. As discussed in Section 4.2.2, a good estimate of the CM, denoted herein by $\mathbf{R}$, is an important ingredient for the SO learning equation. This matrix estimate substitutes the term $(\mathbf{\Omega}_k^H \mathbf{\Omega}_k)^{-1}$ in the GN learning equation, and can be estimated off-line and then kept fixed during the iterative learning process [4], up to some limit [P2, P4, P6, P8] . Typically, if the input signal presents small variations (e.g., changes in allocated resource blocks (RBs), PAPR changes, a reduced signal bandwidth, etc.), this estimate is still quite valid to estimate the learning parameters. However, if the input signal is completely changed (e.g., increased signal bandwidth, changes in the sampling frequency, frequency shifts, etc.), this term needs to be recalculated. This calculation is usually heavy, as a lot of data is typically required to get a good estimate of $\mathbf{R}$. This is where the proposed ICM estimation methods, which drastically reduce this complexity, play their role.

Specifically, five different methods of calculating or manipulating an existing ICM,

are proposed, each one corresponding to a particular scenario. These methods are summarized as follows, and later reviewed in subsequent subsections.

- Estimation of the ICM with the autocorrelation function: Consists of getting a generic estimate of the ICM by using the autocorrelation function of the input data.
- Estimation of the ICM with Bussgang's coefficients: Aims at simplifying the previous method by using Bussgang's decomposition.
- Reduction of the dimensionality of the ICM: Consists of rapidly obtaining the ICM if a reduced model parametrization is desired.
- Estimation of the ICM with a frequency shift: Consists of efficiently shifting the ICM to another target center frequency.
- Recursive estimation of the ICM: Presents a learning rule for the iterative adaptation of the ICM.

It is noted that the first proposed method builds on Reed's theorem [208], [214] for calculating the CM, while the second solution makes use of Bussgang's theorem [66], [214] and other stochastic analysis to provide an approximation for the CM. Since Gaussian process theory is considered, these two methods require Gaussian-distributed input signals. The signals used for verifying the theorems are OFDM signals, which indeed are known to converge to a Gaussian distribution, as the number of subcarriers is generally large. This claim is demonstrated in [P4, P8]. Additionally, these two first methods are especially designed for MP and GMP-like model structures, although they can be modified to fit other system models as well. It is recalled that the associated instantaneous input data vector (i.e., first row of the BF matrix) in a polynomial-based model can be expressed as

$$\begin{aligned}\upsilon^{\mathrm{MP}}[n] = \Big[& x[n] \quad x[n-1] \quad \cdots \quad x[n-M+1] \\ & x[n]|x[n]|^2 \quad x[n-1]|x[n-1]|^2 \quad \cdots \quad x[n-M+1]|x[n-M+1]|^2 \\ & x[n]|x[n]|^{(P-1)} \quad x[n-1]|x[n-1]|^{(P-1)} \quad \cdots \quad x[n-M+1]|x[n-M+1]|^{(P-1)} \Big]^T,\end{aligned} \tag{4.25}$$

for a generic polynomial order P and memory M. The remaining methods can be utilized in every processing model structure, such as MP or GMP [193], Volterra

DDR [106], [272], the proposed MP-LUT model itself, or other LUT models [130], [189].

The computational complexity of each of those solutions is studied in the last subsection. To provide a certain context, it is also compared to the complexity required to calculate the ICM in the classical way. Finally, it is noted that the contents of this section are based on the journal publications [P2, P4] . To be best of the Author's knowledge, these methods have not been applied to closed-loop SO DPD learning applications, thus this combination poses, in part, the novelty of this section.

4.4.1 Estimating the ICM with the Autocorrelation Function

The first proposed method to estimate the ICM is presented herein. In contrast to the following Bussgang's approximation, this approach calculates the proper autocorrelation values by using analytical Gaussian moments theorems, thus a more exact ICM is obtained, with the cost of a slightly increased complexity [P4]. This solution essentially outputs a pre-computed set of coefficients which are dependent on the autocorrelation function of the input data, thus this is the only parameter that is required at the time of calculating the ICM. This method greatly alleviates the computational complexity of calculating this term, as reviewed in the following sections.

The proposed solution is then built as follows. The first step to take is to express the CM exclusively as a function of the autocorrelation function of the input signal. To this end, Reed's Gaussian moment theorem [214] can be used, which states that a pth order central product moment is non-zero if p is even, and can be specifically obtained as the sum of products of covariances. Formally, this is expressed as

$$\begin{aligned}\mathrm{E}\{x^*[n]x^*[n-1]\cdots x^*[n-k+1]x[n]x[n-1]\cdots x[n-k+1]\} = \\ \sum_{\zeta}\mathrm{E}\{x^*[n-\zeta]x[n]\}\,\mathrm{E}\{x^*[n-\zeta]x[n-1]\}\cdots\mathrm{E}\{x^*[n-\zeta]x[n-k+1]\},\end{aligned} \tag{4.26}$$

where ζ represents a permutation of the set $\{0,1,\cdots,(k-1)\}$. All terms in the CM have an even-order polynomial degree, thus Reed's theorem will produce a non-zero

value. Note that terms of the form $\mathrm{E}\{|x[n]|^{2p}\}$ can be alternatively obtained as [228]

$$\mathrm{E}\{|x[n]|^{2p}\} = p!\mathrm{E}\{|x[n]|^2\}^p = p!\sigma_x^{2p}, \qquad p = 1,\ 2,\ \cdots,\ P, \tag{4.27}$$

which may simplify their calculation. Specific examples of how the Reed theorem is applied can be found in [P4], where a fourth and a six-order product moment is developed. It is recommended to program the expressions presented in (4.26) and (4.27) in order to recursively obtain other higher-order terms in the CM. Once each individual term is obtained with Reed's theorem, the resulting CM can be expressed with a specific sub-matrix structure, which will ease its inversion. The overall size of this block-matrix is $\lceil \frac{P}{2} \rceil \times \lceil \frac{P}{2} \rceil$, and the size of each of the sub-matrices is $M \times M$. Formally, such a matrix reads

$$\mathbf{R} \in \mathbb{C}^{\lceil \frac{P}{2} \rceil \times \lceil \frac{P}{2} \rceil} = \begin{bmatrix} \mathbf{R}_{1,1} & \mathbf{R}_{1,M+1} & \cdots & \mathbf{R}_{1,(\lceil \frac{P}{2} \rceil - 1)M+1} \\ \mathbf{R}_{1,M+1} & \mathbf{R}_{M+1,M+1} & \cdots & \mathbf{R}_{M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{1,(\lceil \frac{P}{2} \rceil - 1)M+1} & \mathbf{R}_{M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} & \cdots & \mathbf{R}_{(\lceil \frac{P}{2} \rceil - 1)M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} \end{bmatrix}, \tag{4.28}$$

where each sub-matrix, $\mathbf{R}_{k,l}$, is of the size $M \times M$, and can be defined as

$$\mathbf{R}_{k,l} \in \mathbb{C}^{\lceil M \times M \rceil} = \begin{bmatrix} \mathrm{E}\{\boldsymbol{\upsilon}_k[n]\boldsymbol{\upsilon}_l^*[n]\} & \cdots & \mathrm{E}\{\boldsymbol{\upsilon}_k[n]\boldsymbol{\upsilon}_{l+M-1}^*[n]\} \\ \vdots & \ddots & \vdots \\ \mathrm{E}\{\boldsymbol{\upsilon}_{k+M-1}[n]\boldsymbol{\upsilon}_l^*[n]\} & \cdots & \mathrm{E}\{\boldsymbol{\upsilon}_{k+M-1}[n]\boldsymbol{\upsilon}_{l+M-1}^*[n]\} \end{bmatrix}. \tag{4.29}$$

Here, each subindex in $\boldsymbol{\upsilon}[n]$ indicates the corresponding element within the instantaneous input data vector. This notation allows for expressing these matrices with any polynomial order, P, and memory depth, M. Additionally, only the calculation of the upper triangular submatrices in $\mathbf{R}$ is needed to build the complete CM. Afterwards, this CM is inverted in a classical way to serve as the input of the SO learning equation.

4.4.2 Approximating the ICM with the Bussgang Theorem

The second studied method approximates the ICM by using the Bussgang theorem [66], [185] and the input data autocorrelation function. This method also provides a set of precomputed coefficients that multiply the autocorrelation function of the input data to build the final ICM [P4]. However, the method configures the final ICM with a specific sub-matrix notation, that allows for efficient Kronecker inversion [240]. This specific configuration can be taken if assuming a specific simplification for the Bussgang method, explained later within this subsection, while the exact approach is presented in the previous subsection. The Kronecker inversion alleviates even more the computational complexity required to calculate the ICM, but provides a non-optimal performance, although still close to other state-of-the-art models, as demonstrated in [P4 ,P8] and also in Chapter 5.

This method is specifically tailored to estimate the ICM in MP-type models, thus it is important to configure the BF matrix exactly as presented in (4.25). It is also important to note that, since Bussgang's theorem and other stochastic Gaussian theory is used in this method, the input signals need to follow a Gaussian distribution [66]. Using signals that differ from a Gaussian distribution will result in non-optimal performance [P4, P8].

In order to build the proposed Bussgang method, the first step is to construct the input data BF matrix exclusively as a function of the input data autocorrelation function. Considering an input signal $x[n]$, its autocorrelation and cross-correlation functions can be mathematically defined as

$$R_{\mathrm{X}}(\tau) = \mathrm{E}\{x[n]x^{*}[n-\tau]\}, \tag{4.30}$$

$$R_{\mathrm{XZ}}(\tau) = \mathrm{E}\{x[n]z^{*}[n-\tau]\}, \tag{4.31}$$

where $z[n-\tau]$ corresponds to the output of a nonlinear operator through which the input signal $x[n]$ is passed, and τ corresponds to an arbitrary sample delay. The second order terms of the CM, which appear in its first line, are equivalent to the expression presented in (4.30), thus they can be directly expressed as a function of

the autocorrelation function. These are simply

$$\mathbf{R}(1,1) = R_{\mathrm{X}}(\tau = 0), \tag{4.32}$$

$$\mathbf{R}(1,2) = R_{\mathrm{X}}(\tau = 1), \tag{4.33}$$

$$\vdots$$

$$\mathbf{R}(1,M) = R_{\mathrm{X}}(\tau = M). \tag{4.34}$$

Higher-order terms in the CM are no longer equivalent to (4.30). These higher order terms (i.e., $x[n]|x[n]|^p,\ p = 2,4,\ldots$) can be expressed with the cross-correlation function defined in (4.31), as a function of each corresponding nonlinear operator $z[n-\tau]$ in each term. To evaluate these higher-order correlation values, Bussgang's theorem can be applied [66]. The Bussgang theorem essentially states that the cross-correlation of any input Gaussian signal with the output signal of an instantaneous nonlinear function can be mathematically expressed as the product of a scaling constant and the autocorrelation of the input signal. Formally, this can be formulated as

$$R_{\mathrm{XZ}}(\tau) = \xi R_{\mathrm{X}}(\tau), \tag{4.35}$$

where ξ is the Bussgang coefficient. A proof of Bussgang's theorem for complex-valued signals can be found in [66]. The Bussgang coefficient can then be obtained as

$$\begin{aligned}\xi &= \frac{E\{f(x[n])x^*[n]\}}{E\{|x|^2\}} = \frac{E\{f(x[n])x^*[n]\}}{\sigma_x^2} \\ &= \frac{1}{\pi\sigma_x^4}\int_{-\infty}^{\infty} x^*[n]f(x[n])e^{-\frac{|x[n]|^2}{\sigma_x^2}}\,dx,\end{aligned} \tag{4.36}$$

where the definition of the complex proper[1] Gaussian probability density function (PDF) has been applied [119], [228], and σ_x^2 is the variance of the input signal $x[n]$. This value is usually smaller than one in order to obtain a small condition number in the CM. Using the expressions in (4.35) and (4.36), the remaining high-order terms in the CM can be obtained, exclusively as a function of $R_{\mathrm{X}}(\tau)$ and the Bussgang

[1]Proper (or circular) variable x: zero mean, and x is not correlated with x^*. This property significantly simplifies the Gaussian PDF and complementary distribution function (CDF) [228], [238].

$$\mathbf{R}=\begin{bmatrix}
\xi_0 R_{\mathrm{X}}(\tau=0) & \xi_0 R_{\mathrm{X}}(\tau=1) & \cdots & \xi_0 R_{\mathrm{X}}(\tau=M) & \xi_1 R_{\mathrm{X}}(\tau=0) & \xi_1 R_{\mathrm{X}}(\tau=1) & \cdots & \xi_1 R_{\mathrm{X}}(\tau=M) \\
\xi_0 R^{*}_{\mathrm{X}}(\tau=1) & \xi_0 R_{\mathrm{X}}(\tau=0) & \cdots & \xi_0 R_{\mathrm{X}}(\tau=M-1) & \xi_1 R^{*}_{\mathrm{X}}(\tau=1) & \xi_1 R_{\mathrm{X}}(\tau=0) & \cdots & \xi_1 R_{\mathrm{X}}(\tau=M-1) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\xi_0 R^{*}_{\mathrm{X}}(\tau=M) & \xi_0 R^{*}_{\mathrm{X}}(\tau=M-1) & \cdots & \xi_0 R_{\mathrm{X}}(\tau=0) & \xi_1 R^{*}_{\mathrm{X}}(\tau=M) & \xi_1 R^{*}_{\mathrm{X}}(\tau=M-1) & \cdots & \xi_1 R_{\mathrm{X}}(\tau=0) \\
\xi_1 R_{\mathrm{X}}(\tau=0) & \xi_1 R_{\mathrm{X}}(\tau=1) & \cdots & \xi_1 R_{\mathrm{X}}(\tau=M) & \xi_2 R_{\mathrm{X}}(\tau=0) & \xi_2 R_{\mathrm{X}}(\tau=1) & \cdots & \xi_2 R_{\mathrm{X}}(\tau=M) \\
\xi_1 R^{*}_{\mathrm{X}}(\tau=1) & \xi_1 R_{\mathrm{X}}(\tau=0) & \cdots & \xi_1 R_{\mathrm{X}}(\tau=M-1) & \xi_2 R^{*}_{\mathrm{X}}(\tau=1) & \xi_2 R_{\mathrm{X}}(\tau=0) & \cdots & \xi_2 R_{\mathrm{X}}(\tau=M-1) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\xi_1 R^{*}_{\mathrm{X}}(\tau=M) & \xi_1 R^{*}_{\mathrm{X}}(\tau=M-1) & \cdots & \xi_1 R_{\mathrm{X}}(\tau=0) & \xi_2 R^{*}_{\mathrm{X}}(\tau=M) & \xi_2 R^{*}_{\mathrm{X}}(\tau=M-1) & \cdots & \xi_2 R_{\mathrm{X}}(\tau=0)
\end{bmatrix} \tag{4.41}$$

coefficient, as

$$\mathrm{E}\{x[n]x^{*}[n]|x[n]|^{2}\}=\xi_1 R_{\mathrm{X}}(\tau=0), \tag{4.37}$$

$$\mathrm{E}\{x[n]x^{*}[n-1]|x[n-1]|^{2}\}=\xi_1 R_{\mathrm{X}}(\tau=1), \tag{4.38}$$

$$\mathrm{E}\{x[n]x^{*}[n]|x[n]|^{4}\}=\xi_2 R_{\mathrm{X}}(\tau=0), \tag{4.39}$$

$$\mathrm{E}\{x[n]x^{*}[n-1]|x[n-1]|^{4}\}=\xi_2 R_{\mathrm{X}}(\tau=1), \tag{4.40}$$

$$\vdots$$

where ξ_0, ξ_1, $\cdots$ represent the set of Bussgang coefficients, which can be obtained through (4.36). Now, the second-order and the higher-order terms in the CM have been properly addressed, and are expressed as a function of only the autocorrelation function of the input data and the Bussgang coefficients. A concrete example is presented in (4.41), which shows the CM structure with any value of M, $P=2$, and assuming that $\xi_0=1$. Further analysis of (4.41) reveals that it can be configured with a specific $M\times M$ submatrix notation, that will allow for efficient Kronecker inversion [232], [240]. Hence, the CM can be divided into sub-blocks of size $M\times M$

that contains a specific Bussgang coefficient, and are propagated through it as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} \\ \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{R}_3 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+1} \\ \mathbf{R}_2 & \mathbf{R}_3 & \mathbf{R}_4 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+1} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+2} & \cdots & \mathbf{R}_{2\lfloor \frac{P}{2} \rfloor} \end{bmatrix}, \tag{4.42}$$

where each Hermitian Toeplitz sub-block can be defined as

$$\mathbf{R}_k \in \mathbb{C}^{M \times M} = \xi_k \begin{bmatrix} R_{\mathrm{X}}(\tau=0) & R_{\mathrm{X}}(\tau=1) & \cdots & R_{\mathrm{X}}(\tau=M) \\ R^*_{\mathrm{X}}(\tau=1) & R_X(\tau=0) & \cdots & R_{\mathrm{X}}(\tau=M-1) \\ \vdots & \vdots & \ddots & \vdots \\ R^*_{\mathrm{X}}(\tau=M) & R^*_{\mathrm{X}}(\tau=M-1) & \cdots & R_{\mathrm{X}}(\tau=0) \end{bmatrix}. \tag{4.43}$$

Using (4.43), the covariance matrix in (4.42) can also be expressed as a Kronecker product as [240]

$$\mathbf{R} = \Xi \otimes \mathbf{R}_0, \tag{4.44}$$

where $\Xi \in \mathbb{R}^{\lceil \frac{P}{2} \rceil \times \lceil \frac{P}{2} \rceil}$ represents a matrix containing the complete set of Bussgang coefficients. Finally, the ICM can be obtained by applying the Kronecker inverse as

$$\mathbf{R}^{-1} = \Xi^{-1} \otimes \mathbf{R}_0^{-1}. \tag{4.45}$$

Note that only the first row and the last column of (4.42) need to be obtained in order to estimate the final ICM. It is also noted that other low-complexity inversion methods, such as the block-recursive inverse algorithm [56], can be used to efficiently calculate the matrix inverse of (4.42). However, after careful analysis, it is concluded that the Kronecker approach is the least complex of the two. The complexity of this method is further analyzed at the end of this chapter.

4.4.3 Reducing the Dimensionality of the ICM

The third proposed method aims at avoiding the recalculation of the ICM if a different model parametrization is required (e.g., if the polynomial order in an MP model is reduced, or the number of memory taps is decreased). The proposed method achieves so by directly removing corresponding rows and columns of the original ICM, such that the resulting matrix contains only the entries corresponding to the desired parametrization. It is noted that the coefficients are directly removed from the matrix inverse, thus the process is not straightforward.

The proposed method can then be implemented as follows. Let us denote the original ICM as $\tilde{\mathbf{Q}}^{-1} \in \mathbb{C}^{\tilde{C}\times\tilde{C}} = (\tilde{\mathbf{R}}^{*})^{-1}$, obtained from an arbitrary set of BF, $\tilde{\mathbf{\Omega}}$. All of these matrices are considered known, and they are utilized to obtain the reduced set $\mathbf{Q}^{-1} \in \mathbb{C}^{C\times C} = (\mathbf{R}^{*})^{-1}$, containing only the target entries. The original and reduced BF sets are related as $\tilde{\mathbf{\Omega}} = [\mathbf{\Omega} \quad \mathbf{v}]$, where $\mathbf{v} \in \mathbb{C}^{K^{\text{it}}\times 1}$ corresponds to a column in $\tilde{\mathbf{\Omega}}$ that needs to be removed. The original CM can be expressed as

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{\Omega}}^{H}\tilde{\mathbf{\Omega}} = \begin{bmatrix} \mathbf{\Omega}^{H} \\ \mathbf{v}^{H} \end{bmatrix} \begin{bmatrix} \mathbf{\Omega} & \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}^{H}\mathbf{\Omega} & \mathbf{\Omega}^{H}\mathbf{v} \\ \mathbf{v}^{H}\mathbf{\Omega} & \mathbf{v}^{H}\mathbf{v} \end{bmatrix}, \tag{4.46}$$

where the Schur Complement [203] can be applied to obtain its block inverse. It reads

$$\tilde{\mathbf{Q}}^{-1} = \left[\begin{array}{c|c} \mathbf{A}^{-1} & \mathbf{B}^{-1} \\ \hline \mathbf{C}^{-1} & \mathbf{D}^{-1} \end{array}\right], \tag{4.47}$$

where $\mathbf{A} \in \mathbb{C}^{(\tilde{C}-1)\times(\tilde{C}-1)}$, $\mathbf{B} \in \mathbb{C}^{(\tilde{C}-1)\times 1}$, $\mathbf{C} \in \mathbb{C}^{1\times(\tilde{C}-1)}$, and $\mathbf{D} \in \mathbb{C}^{1\times 1}$. These terms,

in turn, read [203]

$$\mathbf{A}^{-1} = (\mathbf{\Omega}^H\mathbf{\Omega})^{-1} + \mathbf{D}^{-1}(\mathbf{\Omega}^H\mathbf{\Omega})^{-1}\mathbf{\Omega}^H\mathbf{v}\mathbf{v}^H\mathbf{\Omega}\Big((\mathbf{\Omega}^H\mathbf{\Omega})^{-1}\Big)^H, \tag{4.48}$$

$$\mathbf{B}^{-1} = -\mathbf{D}^{-1}(\mathbf{\Omega}^H\mathbf{\Omega})^{-1}\mathbf{\Omega}^H\mathbf{v}, \tag{4.49}$$

$$\mathbf{C}^{-1} = -\mathbf{D}^{-1}\mathbf{v}^H\mathbf{\Omega}\Big((\mathbf{\Omega}^H\mathbf{\Omega})^{-1}\Big)^H, \tag{4.50}$$

$$\mathbf{D}^{-1} = \Big(\mathbf{v}^H\mathbf{v} - \mathbf{v}^H\mathbf{\Omega}(\mathbf{\Omega}^H\mathbf{\Omega})^{-1}\mathbf{\Omega}^H\mathbf{v}\Big)^{-1}. \tag{4.51}$$

Note that the term $(\mathbf{\Omega}^H\mathbf{\Omega})^{-1}$ appearing in $\mathbf{A}^{-1}$ directly corresponds to the desired ICM subset, thus it can be finally obtained as

$$\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1} = (\mathbf{\Omega}^H\mathbf{\Omega})^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1}\Big(\frac{\mathbf{B}^{-1}}{\mathbf{D}^{-1}}\Big)^H. \tag{4.52}$$

It is noted that removing one column from the original BF matrix implies removing one row and one column from the associated matrix inverse $\tilde{\mathbf{Q}}^{-1}$. The target row and column to be removed are required to be in the last row/column of the ICM, thus this method can be implemented iteratively to remove any arbitrary number of rows/columns.

4.4.4 Frequency Shifting the ICM

The fourth method consists of efficiently shifting the CM in order to obtain a replica that can be used in another target center frequency. If the input signal is frequency upshifted or downshifted, its input statistical content will be different from the original, and so its CM will also vary [176], [241]. Thereby, in order to obtain an optimal performance, the CM and ICM need to be estimated again by some means. A novel approach that directly applies a frequency shift to obtain the new ICM matrix, instead of crudely recalculating it again, is presented within the next lines. For simplicity, this method is presented in the context of MP processing, reviewed in Section 2.2, but can be applied for any other model requiring the ICM calculation in the SO algorithm. The method is built as follows.

Let $\tilde{\mathbf{Q}}^{-1} = (\tilde{\mathbf{R}}^*)^{-1}$ be the original ICM before applying the frequency shift, which has been obtained from a signal $\tilde{x}[n]$ and BF matrix $\tilde{\mathbf{\Omega}}$. Let $\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1}$ the resulting ICM the method aims at estimating, corresponding to a signal $x[n]$ and BF ma-

$$\mathbf{\Omega}^1 = \Big[x[n]e^{j\phi\omega n} \cdots x[n]|x[n]|^{(P-1)}e^{j\phi\omega n} \quad x[n-1]e^{j\phi\omega(n-1)} \cdots x[n-1]|x[n-1]|^{(P-1)}e^{j\phi\omega(n-1)} \cdots$$
$$\cdots x[n-M+1]e^{j\phi\omega(n-M+1)} \cdots x[n-M+1]|x[n-M+1]|^{(P-1)}e^{j\phi\omega(n-M+1)} \Big] \quad (4.54)$$

---

trix $\mathbf{\Omega}$. An arbitrary frequency shift of ϕf, in Hz, can be mathematically expressed in the time domain as the multiplication with a phase term [198]. Considering $\tilde{x}[n]$ as the input signal, and $x[n]$ as the resulting shifted signal, the frequency shift reads

$$x[n] = e^{j\phi\omega n}\tilde{x}[n], \quad (4.53)$$

where $\phi\omega = (2\pi f)/f_s$ is the normalized angular frequency shift, and f_s is the sampling frequency. Having the frequency shift defined, the shifted input data BF matrix, $\mathbf{\Omega}$, can be expressed exclusively as a function of the unshifted signal $\tilde{x}[n]$ and the frequency shift $\phi\omega$. This formulation can be obtained by substituting (4.53) in the input BF matrix, obtained through the formulation presented in Section 2.2. The first row of the resulting BF matrix is presented in (4.54). As a result of defining the BF matrix in such a way, the entries of the conjugated ICM, $\mathbf{Q}$, can be expressed as a function of the unshifted ICM, $\tilde{\mathbf{Q}}$, and an specific weight which depends on $\phi\omega$. According to the matrix structure of (4.54) and the associated BF matrix in Section 2.2, the weights are dependent on the different memory terms appearing in each matrix, and can be divided in zero-sample-delay terms, one-sample-delay terms, and, in general, m-sample-delay terms, as

$$\begin{aligned}
&0\text{-sample-delayterms: } \mathbf{Q}(i,j) = e^{-j\phi\omega n}e^{j\phi\omega n}\tilde{\mathbf{Q}}(i,j) = \tilde{\mathbf{Q}}(i,j),\\
&1\text{-sample-delayterms: } \mathbf{Q}(i,j) = e^{-j\phi\omega n}e^{j\phi\omega(n-1)}\tilde{\mathbf{Q}}(i,j) = e^{-j\phi\omega 1}\tilde{\mathbf{Q}}(i,j),\\
&\qquad\qquad\vdots\\
&m\text{-sample-delayterms: } \mathbf{Q}(i,j) = \cdots = e^{-j\phi\omega m}\tilde{\mathbf{Q}}(i,j).
\end{aligned} \quad (4.55)$$

The zero-sample-delay terms are usually around the main diagonal of $\tilde{\mathbf{Q}}$, and larger sample-delay terms are scattered through it. The final structure of the ICM and an specific example of its configuration can be found in [P4]. The key insight until this point is that the resulting shifted ICM only depends on the original ICM, $\tilde{\mathbf{Q}}$, and the frequency shift $\phi\omega$. Thereby, the final shifted ICM can be simply obtained

by properly multiplying the original ICM with the proper frequency shift weights, corresponding to different delayed terms in the input BF matrices. One method to do so is to express the shifted CM as

$$\mathbf{Q} = \mathbf{P}^H \tilde{\mathbf{Q}} \mathbf{P}, \tag{4.56}$$

where the transformation matrix $\mathbf{P} \in \mathbb{C}^{C \times C}$ applies the correct weights for each memory term in the BF matrix. This matrix will vary according to the order of chosen BFs. In this case, it is defined as

$$\mathbf{P} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & e^{j\phi\omega m} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & e^{j\phi\omega m} \end{bmatrix}. \tag{4.57}$$

Finally, thanks to the nomenclature chosen in (4.56), the CM can be simply inverted as

$$\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1} = \mathbf{P}^{-1} \tilde{\mathbf{Q}}^{-1} \mathbf{P}^{H(-1)}. \tag{4.58}$$

The overall process drastically reduces the computational complexity of estimating the ICM, in contrast to performing its whole recalculation. This is demonstrated in [P4] and also in the following Chapter 5. It is finally noted that the number of parameters, C, in DPD applications is not usually large, thus the matrix inversion appearing in (4.58) does not constitute, in general, a heavy term when estimating the ICM.

4.4.5 Recursively Estimating the ICM

The last proposed method aims at recursively estimating the ICM in target learning iterations. This method is especially useful when the input data signal changes dynamically, and thus the ICM can be continuously estimated, providing optimal

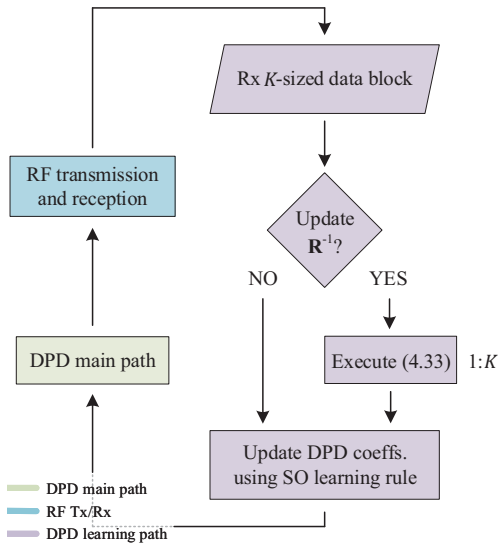


Figure 4.2 Conceptual flow diagram of the recursive ICM estimation method. Note that the learning rule presented in (4.60) does not need to be executed at every learning iteration, and a reduced estimation sample data set can be selected for its update.

performance at every time instant. The derived learning rule for the estimation of the ICM does not need to be executed in every learning iteration, and even a reduced estimation sample data set can be considered, per iteration, to estimate the ICM. A conceptual data flow of the proposed algorithm is depicted in Fig. 4.2, showing the idea of the overall process.

The CM can be first estimated as

$$\mathbf{R}_{n+1} = \frac{n-1}{n}\mathbf{R}_n + \frac{1}{n}\mathbf{x}_n^{\mathrm{BF}}\mathbf{x}_n^{\mathrm{BF},H}, \quad n = 1,2,\cdots,K^{\mathrm{it}}, \tag{4.59}$$

where n represents the time index, $\mathbf{x}_n^{\mathrm{BF}}$ is the nth row of the input data BF matrix, $\mathbf{\Omega}$, and $\mathbf{R}_0$ can be initialized as a zero matrix in the first learning iteration. Once having the direct adaptation of the CM, its inverse can easily be obtained by applying the matrix inversion lemma [31], [203], as

$$\mathbf{R}_{n+1}^{-1} = \frac{n}{n-1}\left(\mathbf{R}_n^{-1} - \frac{1}{n-1+\mathbf{x}_n^{\mathrm{BF},H}\mathbf{R}_n^{-1}\mathbf{x}_n^{\mathrm{BF}}}\mathbf{R}_n^{-1}\mathbf{x}_n^{\mathrm{BF}}\mathbf{x}_n^{\mathrm{BF},H}\mathbf{R}_n^{-1}\right), \qquad n = 2,3,\cdots,K^{\mathrm{it}}, \tag{4.60}$$

where, again, $\mathbf{R}_0$ can be initialized as a zero matrix in the first learning iteration. Since the ICM does not need to be updated in every learning iteration, this method poses a clear advantage compare to GN, which requires the calculation of the term $(\mathbf{\Omega}_k^H \mathbf{\Omega})^{-1}$ in every learning data block. Overall, this method allows for significant complexity reductions in the learning path, while achieving a close performance to the reference GN solution.

4.4.6 Complexity Analysis

In this section, a complexity comparison of the earlier presented ICM estimation methods is analyzed. Also, the complexity of calculating the ICM in a classical way is shown, for reference. This complexity analysis is done in terms of real multiplications per K^{it}-sized learning iteration, and left indicated as a function of the model parameters, while concrete numerical examples are later detailed in [P4, P8] . The complexity analyses are done by following the exact processing steps reviewed in Sections 4.4.1 to 4.4.5. When considering the third ICM estimation method, presented in Section 4.4.3, the parameter l^{ICM} represents the number of rows and columns that need to be removed from the original ICM. When considering the recursive ICM estimation method, presented in Section 4.4.5, it is assumed that the learning equation for the ICM in (4.60) is executed in only τ^{ICM} samples within the complete K^{it}-sized learning block. This is a plausible assumption since the ICM generally converges fast, and thus not that many iterations are needed. Note that if $\tau^{\text{ICM}} = K^{\text{it}}$, the update in (4.60) would be executed in all samples of the K^{it}-sized learning data block.

The complete set of complexity expressions corresponding to the presented ICM estimation methods are presented in Table 4.4, as a function of the modeling parameters, and following the assumptions explained above. The table also gathers a numerical example, when the ICM parametrization is chosen as $P = 9$, $M = 4$, $C = 20$, $K^{\text{it}} = 20.000$, $l^{\text{ICM}} = 5$, and $\tau^{\text{ICM}} = K^{\text{ICM}}/4$. This parametrization is similar to the one chosen to verify the various ICM estimation methods. More specific numerical comparisons and examples are further reviewed in [P4, P8] , where also the complexity-performance trade-offs of the proposed models are reported.

Table 4.4 Computational complexity of the various ICM estimation methods reviewed in Sections 4.4.1 to 4.4.5, in terms of real multiplications per K^{it}-sized learning block, and as a function of the model parameters. The classical ICM calculation complexity is also included for reference. Also, a numerical example is presented, where $P = 9$, $M = 4$, $C = 20$, $K^{\text{it}} = 20.000$, $l^{\text{ICM}} = 5$, and $\tau^{\text{ICM}} = K^{\text{ICM}}/4$

ICM estimation method	Symbolic real mul. per K^{it}-sized learning block	Numeric real mul. to estimate ICM
Classical ICM calculation	$4C^2(C+K)$	32.032.000
Autocorr. ICM est. (Sec. 4.4.1)	$(\lfloor \frac{P}{2} \rfloor^2 + 3\lfloor \frac{P}{2} \rfloor + 2)3M^2 + 4(C^3(M+1)(K^{\text{it}}+1))$	433.460
Bussgang's ICM est. (Sec. 4.4.2)	$M(4M^2 + 6M\lceil \frac{P}{2} \rceil - 3M + 4K^{\text{it}} + 4) + 4(\lceil \frac{P}{2} \rceil + K^{\text{it}} + 1)$	401.208
Dimensionality red. ICM (Sec. 4.4.3)	$4l^{\text{ICM}}C(C-1)$	7.600
Frequency shifting ICM (Sec. 4.4.4)	$(C^2 - C)/2$	190
Recursive est. of ICM (Sec. 4.4.5)	$\tau^{\text{ICM}}C(18C + 4 + 2div)$	36.400.000

5 EXPERIMENTAL VERIFICATION

THIS chapter provides several RF measurement examples with the aim of verifying the proposed methods reviewed within this thesis. The Chapter is subdivided into two main sections, dealing with DPD- and IBFD DSIC-related applications. Into the case of DPD, the proposed methods serve as the digital predistorter, which, together with their associated learning stages, compose the overall DPD system to test. In the case of DSIC, the proposed methods serve as the digital SI cancellation stage to form a complete IBFD prototype, such that the SI cancellation can be evaluated. In order to provide realistic insight into the performance-complexity trade-offs of the models, complete real-life measurement environments are adopted in all cases.

The measurement results presented in this chapter are based on the publications contained in [P1-P8], and further complemented with the works presented in [146], [155], [200].

5.1 RF Verification in Digital Predistortion

This section deals with the verification of the proposed DPD systems. First, the adopted experimental RF measurement setups are described, and then the proposed cascaded and linear-in-parameters models are tested.

5.1.1 Deployed RF Measurement Setups

This section reviews the two main measurement setups deployed to carry out the DPD measurements. The first setup consists of an FR-1 measurement environment, where sub-6 GHz PAs are measured. Specifically, the RF testing example shown in this thesis is performed with a 5G NR Band n78 medium-range BS PA. The second setup conforms to an FR-2 measurement environment, with a specific mmW center

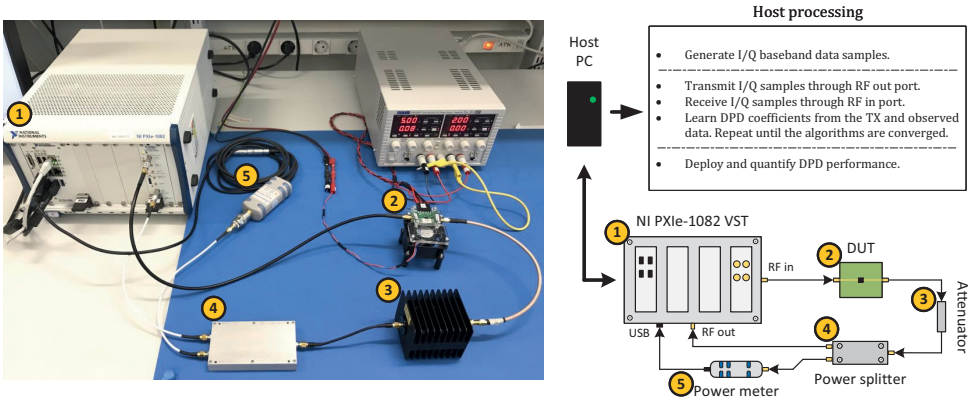


Figure 5.1 Adopted 5G NR FR-1 RF measurement environment to test and evaluate the proposed DPD methods at sub 6 GHz frequencies.

frequency of 28 GHz. With this configuration, a state-of-the-art active antenna array containing 64 integrated PAs and antenna units is linearized. All FR-2 measurements are carried out OTA. More specific characteristics of each setup are detailed in the following subsections.

FR-1 Measurement Environment

The adopted FR-1 measurement setup utilized to test the sub-6 GHz PA is depicted in Fig 5.1, and it is configured as follows [P1]. First, a National Instruments PXIe-5840 vector signal transceiver (VST) facilitates the baseband I/Q pseudo-random input waveform generation, with a maximum instantaneous bandwidth of 1 GHz. The baseband complex input waveform is generated by MATLAB in the VST itself, and transmitted to the device under test (DUT) through its RF output transmit chain. In this particular example, the DUT is a low-to-medium power Skyworks SKY66293-21 amplification unit, oriented to be used in large-array antenna systems or in medium-range BSs [1]. This particular Band n78 (3300-3800 MHz) amplifier has a gain of +34 dB, and a 1-dB compression point of +31.5 dBm. Additionally, its AM/AM response is shown in the Appendix, for reference. After the PA, the signal is carried out through an additional attenuator, and finally fed back to the VST. The VST module acts also as the observation receiver (ORX), and includes a post-processing environment where the DPD algorithms are executed and the DPD

performance metrics are evaluated. The setup in Fig 5.1 includes an additional power splitter, with one of its outputs connected to a power meter to accurately measure the PA output power. With this setup, the transmit signals are generated with a 4× oversampling factor. Further details regarding the signal generation, sampling frequencies, resource block (RB) allocation, subcarrier spacing (SCS), and other physical layer parameters can be found in [P1].

This setup was extensively used in [P1] to evaluate the DPD performance.

FR-2 Measurement Environment

The second measurement environment, corresponding to OTA FR-2 frequencies, is depicted in Fig 5.2, and is configured as follows. First, a Keysight M8190 arbitrary waveform generator (AWG) outputs the complex baseband I/Q data samples at intermediate frequency (IF) 3.5 GHz, and a N5183B-MXG signal generator with an LO signal at 24.5 GHz and a Marki Microwave T31040 mixer further upconvert the generated I/Q signal to 28 GHz center frequency. The resulting signal is then filtered by a Marki Microwave FB3300 BPF to eliminate the mixer-induced image frequencies. Two preamplifiers, an HMC499LC4 and an Analog Devices HMC1131 are configured before the actual array to provide a sufficiently high power at its input, such that it is operated in a well saturated region. The test PA is a 64-element Anokiwave AWMF-0129 active antenna array, which transmits the signal OTA. The array is mounted on an electrical tripod, which is capable of precisely providing the desired elevation and azimuth angles. In a typical DPD experiment, the array *beamforming angle* is 0°, unless otherwise mentioned. For reference, the AM/AM curve of the antenna array is depicted in the Appendix. The radiated signal is later captured by a horn antenna acting as the ORX, which is well aligned to the beamforming angle of the active array. The captured signal is then downconverted, attenuated, and finally fed to a Keysight DSOS804A oscilloscope which is acting as the actual digitizer. The received digital waveform is then given to a host computer for further post-processing.

It is important to note that the RX OTA horn antenna is used for DPD learning as well as the final OTA measurements assessing the DPD performance. This RX is used also as the ORX because the AWMF-0129 array does not provide the option of installing hardware-based combiners to provide the feedback signal [43], [52], [169], [170]. Thus, the horn antenna ORX is deliberately chosen to mimic such a

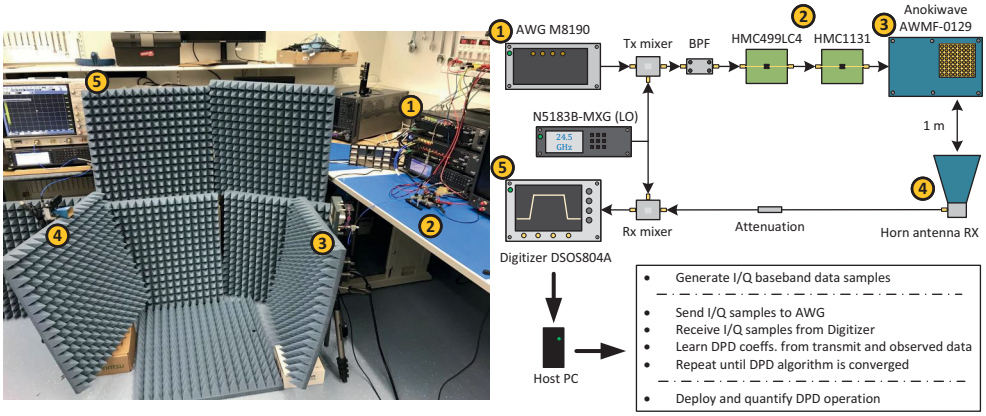


Figure 5.2 Adopted 5G NR FR-2 OTA RF measurement environment to test and evaluate the proposed DPD methods at mmW frequencies.

setup [2], [42]. A detailed study showing the associated OTA to be problematic, such as OTA ORX misalignment and beam-dependence of radiated nonlinear distortion, is presented in [P2].

This setup was used in [P2-P4, P6, P8] to evaluate and test the DPD performance, while [P7] considered a slightly modified version of it.

Finally, it is important to highlight several issues when mmW array setups are considered. First, an array which contains R antennas also contains R PA units, one for each antenna. These PA units are mutually different up to some limit, thus each PA essentially presents its own nonlinear response. The estimated digital predistorter builds on the combined learning path and the combined observed signal, and thus it can only provide a good linearization in the *main beam direction* of the antenna array. At other angles, the beam-pattern of the array itself will keep the OoB distortion levels under specified limits [2], [17], [85], [191]. Second, the nonlinear characteristics of the array have been observed to be beam-dependent [6], [42], stemming from the load modulation phenomena of the PAs, caused by antenna coupling [13]. This indicates that the optimal DPD estimate will depend on the array *main beam direction*, and thus the overall digital predistorter must take this into account [P2]. Plausible solutions are those capable of adapting the DPD coefficients as the beam is steered, such as real-time tracking and fast adaptive DPD learning [45], [180], [207]. This issue especially calls for low-complexity learning algorithms, capable of adapting the DPD estimate in a simple manner [118], [178]. Third, very wide

channel bandwidths, more difficult impedance matching, and calibration challenges already appear in mmW frequencies, compared to other lower frequencies [42], [67], [85], [151]. This essentially translates to a higher frequency selectivity of the TX system, and also that of the nonlinear distortion. As an example, distinct frequency selectivity could appear in different components of the TX IF parts and the PA, and the predistorter may interpret this effect as part of the PA response, hence overestimating its modeling and requiring an unnecessarily high level of parametrization and complexity.

It is finally noted that all the tested input multicarrier waveforms are 3GPP Release 15 compliant [1], unless otherwise mentioned. When using this setup, the signals are generated with a $5\times$ oversampling factor. Further details regarding the signal generation, sampling frequencies, resource block (RB) allocation, subcarrier spacing (SCS), and other physical layer parameters can be found in [P1-P8].

5.1.2 RF Verification of Spline-based Cascaded Models

In this section, RF verification examples are presented to show the effectiveness of the cascaded-based DPD models in a real-life scenario. All cascaded models are estimated with the ILA, as reviewed in Section 2.2. Specifically, two different examples are presented. The first one analyzes the spectral linearization performance of the cascaded models, when considering FR-1 and FR-2 scenarios. In both cases, it is shown how the newly introduced 3GPP Release 15 specification requirements are fulfilled in all cases [1]. The second experiments shows the DPD performance as a function of the effective isotropic radiated power (EIRP) of the array, and addresses several issues appearing in mmW OTA systems.

ACLR and TRP ACLR Limit Violation Example

In this linearization example, the DPD performance obtained with the FR-1 and FR-2 setups is presented. In the results, the classical MP model is also shown, for reference. First, the FR-1 setup is considered, with the previously described Skyworks SKY66293-21 acting as the PA module. Since the radiated power of this PA corresponds to a Medium-Range BS [P3], the absolute basic ACLR limit metric is used to evaluate the performance [1]. The overall linearization results are depicted in

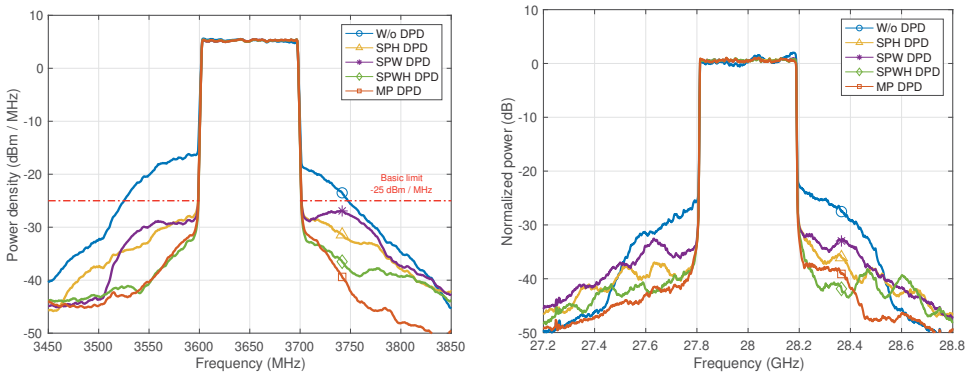
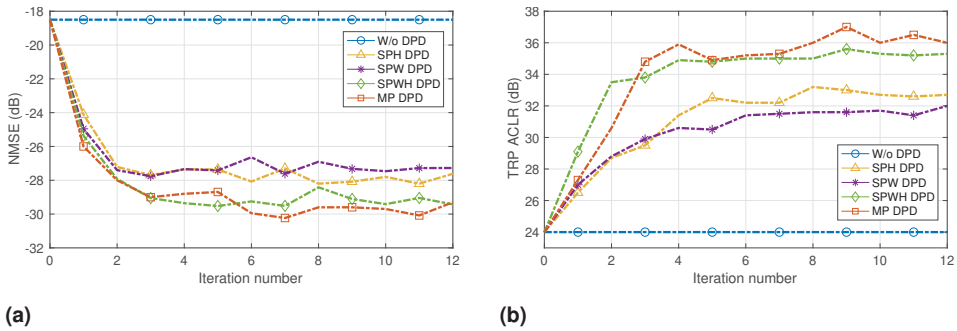


Figure 5.3 5G NR linearization example of the proposed cascaded models in (a) FR-1 band n78 100 MHz measurement environment, using a medium-range BS PA operating at +26 dBm and (b) FR-2 band n257 400 MHz measurement environment, using the active antenna array operating at an EIRP of +42 dBm.

Fig 5.3a. As can be seen from the figure, the OoB nonlinear distortion generated by the PA is clearly violating the specified basic ACLR limit (-25 dBm/MHz) when no DPD is used. However, when using the proposed cascaded DPD models, the distortion power is well below the specified limit, with a specific margin that goes up to +6 dB in the case of SPWH. This result is expected, since the SPWH model constitutes the richest approach out of the ones presented. Second, the FR-2 setup is deployed, with the Anokiwave AWMF-0129 antenna array as the DUT. The overall linearization example is presented in Fig 5.3b. In this case, the specified FR-2 TRP ACLR limit (+28 dBc) is also violated when no DPD is used, with a value around +25 dBc TRP ACLR. However, when the DPD solutions are applied, the TRP ACLR of the resulting curves is diminished such that it complies with the 3GPP Release 15 specifications in all cases. Note also how the passband distortion is corrected by the DPD solutions, also fulfilling in all cases the required EVM limits, which are studied in the next subsection. As an additional feature of this experiment, the DPD convergence of the different cascaded algorithms is studied and presented in Fig. 5.4. The figure shows the NMSE and TRP ACLR numbers as a function of the iteration index, when considering a block iteration sample size of $K^{\text{it}} = 25$ ksamples. It can be observed that the DPD convergence is fast in general, achieving the steady-state behavior in around 4-5 DPD iterations. This result also confirms that the 3GPP TRP ACLR limit is fulfilled in all cases when the DPD algorithms have converged. Finally, it is



(a) (b)

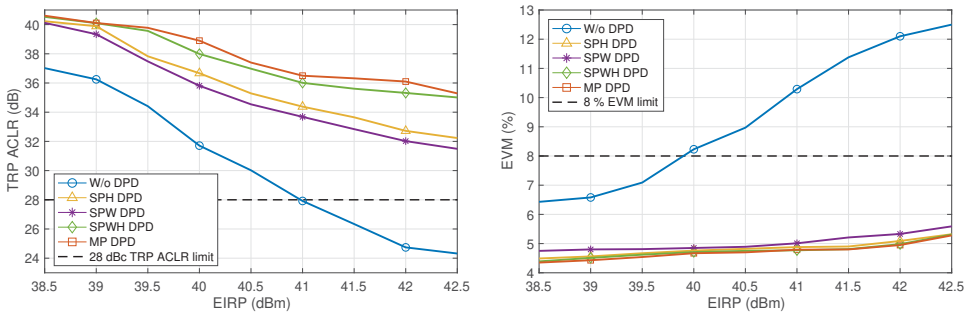
Figure 5.4 5G NR 400 MHz OTA DPD convergence performance in terms of (a) NMSE, and (b) TRP ACLR of the proposed DPD cascaded models as a function of the iteration index, with an EIRP of +42 dBm.

worth mentioning that the main path complexity of the cascaded models is reduced by 68% (SPH), 68% (SPW), and 52% (SPWH), when compared to the classical MP processing, as demonstrated in [P1] and [P5].

DPD Performance vs. EIRP in OTA Systems

One aspect worth highlighting when working with mmW OTA systems is how the DPD algorithms perform as the active array output power varies. To this end, a power sweep measurement example is presented next, showing the TRP ACLR and EVM performance metrics against the EIRP of the array, which ranges from +38.5 dBm to +42.5 dBm. The considered signal bandwidth for this experiment is 400 MHz, while the FR-2 OTA measurement environment is configured exactly as explained in Section 5.1.1. This experiment sheds some light into two main things: i) whether the TRP ACLR or the EVM is the limiting metric of the system, in terms of maximum reachable EIRP, and ii) to test the DPD performance of the developed methods as the array output power increases.

The obtained example results are then presented in Fig. 5.5. From the figure, it can be clearly seen that the limiting metric of the array system is the EVM, which has a value of +8% [1] at an EIRP of already +39.8 dBm, where the TRP ACLR is still under +28 dBc. Nevertheless, when the proposed predistort techniques are applied, both EVM and TRP ACLR performance metrics are kept well under the specified limits at least up to an EIRP of +42.5 dBm, and clearly far beyond, as can be seen from the example figure. These results clearly indicate an output power increase, and thus a power efficiency improvement achievable with the proposed DPD



(a) (b)

Figure 5.5 5G NR 400 MHz OTA linearization performance of the propose DPD methods as a function of the EIRP, in terms of (a) TRP ACLR, and (b) EVM.

solutions, which allow to increase the EIRP to +42.5 dBm and beyond, compared to the maximum +39.8 dBm that would have been possible if no DPD correction was applied. Further power sweep experimentation can be found, for other different DPD models, in [P1-P3].

5.1.3 RF Verification of Spline-based Linear-in-parameters Solutions

In this section, the performance of the several linear-in-parameters solutions presented in Chapter 4 is tested. To this end, several RF measurements examples are presented, which provide proof of the achieved DPD performance. Specifically, a linearization example of the SPMP model is presented, followed by a performance analysis of the sign-based learning algorithms combined with the MP-LUT model, a further DPD performance vs. adopted LUT size comparison, and finally followed by an example of the ICM estimation methods for the SO learning rule.

SPMP ILA Performance Example

In this predistortion example, the linearization performance obtained with the SPMP model is tested, when configured with an ILA as presented in Section 2.2. In this case, the learning procedure is based on the steepest gradient-descent algorithm presented in Section 4.1. The FR-2 measurement setup described in Section 5.1.1 is considered as the test scenario, along with the active antenna array DUT described before. As can be seen from Fig 5.6, the specified FR-2 TRP ACLR requirement

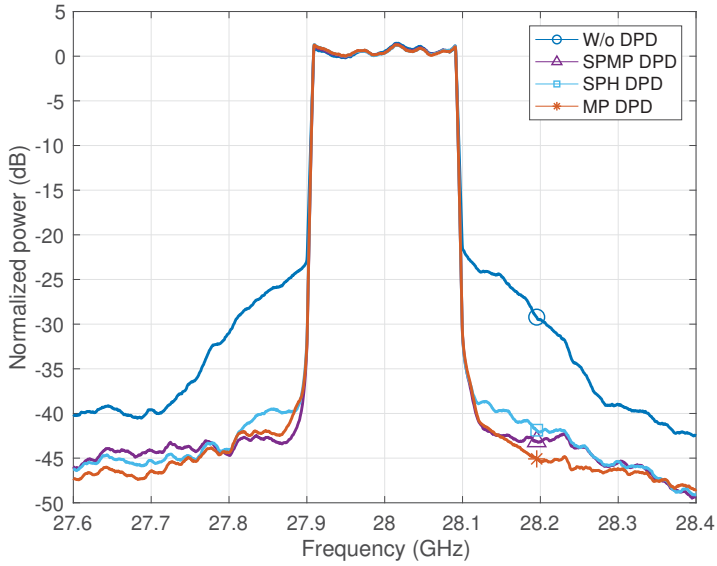
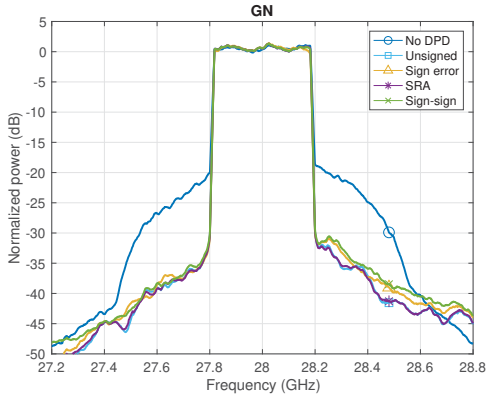


Figure 5.6 A 5G NR linearization example of the proposed SPMP ILA linear-in-parameters model, in an FR-2 band n257 200 MHz measurement environment, using the active antenna array operating at an EIRP of +42.5 dBm.

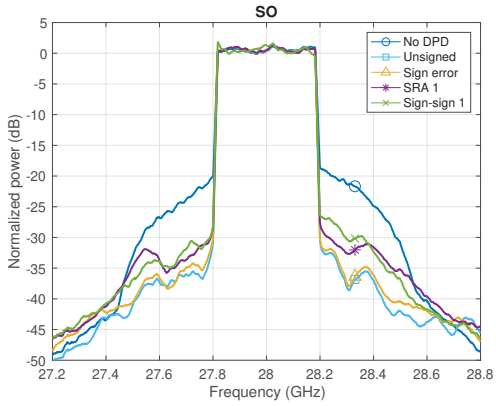
of +28 dBc [1] is violated when no DPD is applied, with a value around +23 dBc. Nevertheless, when the SPMP DPD model is applied, the TRP ACLR is improved up to +38 dBc, which constitutes a consistent margin of +10 dB with respect to the +28 dBc limit [1]. Linearization examples of the previously presented SPH model and an MP model are also presented for reference, when adopting a similar DPD parametrization to make the comparison fair. The computational complexity of the SPMP DPD model is, in turn, reduced by 44% with respect to the classical MP model. A more detailed complexity analysis can be found in [P1].

Signed Closed-loop Algorithms Performance Example

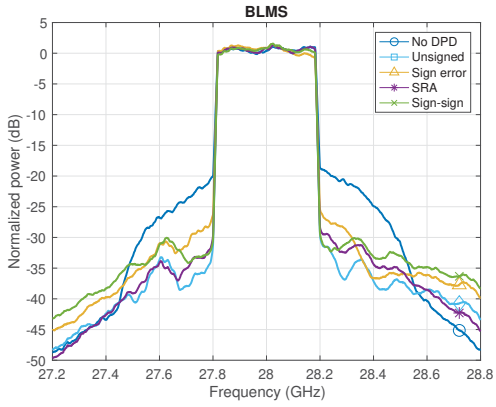
The next linearization example shows the behavior of the sign-based algorithms when combined with the classical GN, SO, and BLMS learning equations. In this example, the mmW FR-2 setup is also adopted, with the active array acting as the DUT nonlinear PA. The results are then presented in Fig. 5.7a, for the three adopted learning equations. As can be seen from all figures, PA linearization is indeed needed, as



(a)



(b)



(c)

Figure 5.7 A 5G NR linearization example of the proposed sign-based algorithms in an FR-2 band n257 400 MHz measurement environment, using the active antenna array operating at an EIRP of +43 dBm, with (a) GN learning method, (b) SO learning method, and (c) BLMS learning method.

the spectral curves without DPD do not comply with the +28 dBc TRP ACLR limit specified in [1]. The TRP ACLR is however diminished under the specified limit when applying all the sign-based learning solutions. In the first case, the GN, the DPD performance of all considered sign-based versions is similar. With the second considered learning rule, the SO, the difference in terms of performance between the sign-based versions is more notable, as further signing its terms already constitutes a more crude approximation. This is observed especially in the cases of the SO sign and SO sign-sign, where the linearization performance drops around +4 dB, when compared to the unsigned classical solution. Finally, with the BLMS solution, the performance is generally degraded, and the convergence speed is also slower [P4, P6], when compared to the GN or SO learning solutions, although its complexity is also lighter. Specific experiments and analyses targeting the DPD convergence speed of the proposed algorithms can be found in [P4,P6]. In general, the computational complexity is, in turn, reduced by up to 50% in the case of the GN, and by more than 99% in the case of the SO and BLMS, as demonstrated in [P2] and [P6].

All in all, the GN presents a robust solution, but often involves an increased complexity. The SO solution presents quite a good approximation, while the complexity is reduced. Finally, the BLMS is a very simple solution, but the performance and convergence may suffer in some scenarios.

DPD Performance vs. LUT Size Comparison Example

One aspect worth noting when working with LUT-based systems is how the considered LUT size affects the achievable DPD performance. This section presents a measurement example showing the DPD behavior against the considered LUT size, when adopting the SO learning rule and its different signed variants. In this FR-2 OTA example, an effective signal bandwidth of 400 MHz and an EIRP of +43 dBm are considered, while the measurement setup is configured as explained in Section 5.1.1. Similar comparisons made with the GN and BLMS learning approaches can be found in [P2].

The example results are then presented in Fig. 5.8, showing the achievable TRP ACLR numbers as a function of the LUT size. The obtained curve when no DPD is applied and the +28 dBc TRP ACLR performance limit [1] are also shown for reference. As can be seen from the figure, the various DPD curves pose a clear ascending pattern, showing an improved linearization performance as the LUT size increases.

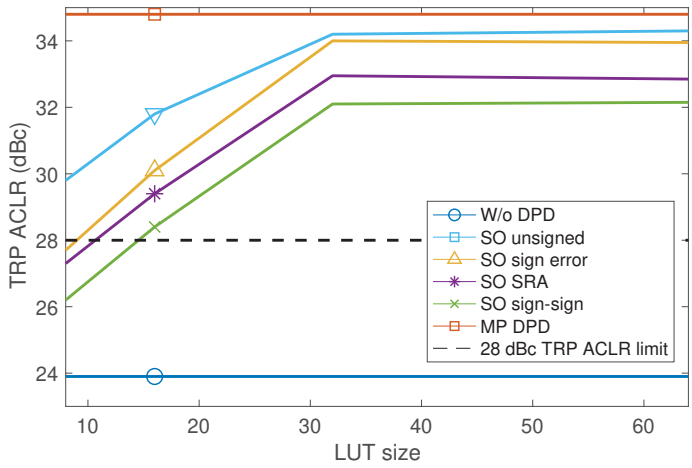


Figure 5.8 Comparison of the achievable TRP ACLR against the adopted LUT size in the SO learning rule and its different signed variants, considering a signal bandwidth of 400 MHz and at an EIRP of +43 dBm.

This effect is seen until the LUT size reaches 32 elements, where the achievable DPD performance essentially saturates. Hence, a 32-point LUT seems to be the optimal choice in this particular example, in terms of linearization performance and required resources. To see a reduced performance when the LUT size is low is naturally expected, since a very low number of control points in the LUT are not sufficient to approximate and invert the effective PA nonlinearity. It is finally noted that the +28 dBc TRP ACLR performance metric is fulfilled in all cases when the LUT size is greater than 8 points. This figure also reconfirms the potential of the proposed sign-based solutions, which are in all cases within the TRP specified limits, regardless of their much lower involved computational complexity.

Another important aspect of the developed methods is their convergence speed. In general, the convergence speed of the unsigned learning equation (i.e., the learning equation when no sign algorithm has been applied) is faster, reaching the steady-state sooner. When applying the different sign-based versions of the learning equations, the convergence speed is slightly decreased, since some information is inevitably lost when signing different terms in them. In all cases, the proposed models are very stable after convergence, and satisfy the 3GPP requirements [1]. Detailed convergence and stability analyses of the proposed methods can be found in [P2, P6].

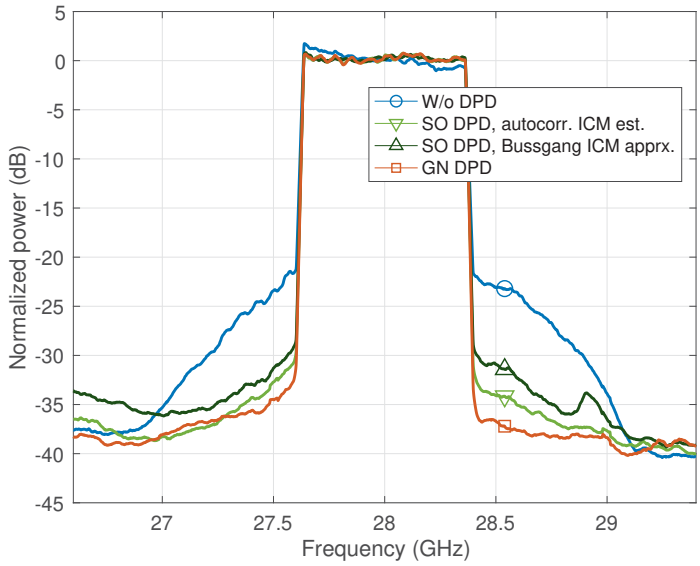


Figure 5.9 A 5G NR linearization example of the proposed Bussgang and autocorrelation ICM estimation methods, combined with the SO learning equation, in an FR-2 band n257 800 MHz measurement environment, and using the active antenna array operating at an EIRP of +40.5 dBm.

5.1.4 RF Verification of the ICM Estimation Methods

The DPD verification is concluded with the testing of the ICM estimation methods. In this section, a linearization example of estimating the ICM with the autocorrelation function and the Bussgang coefficients (Section 4.4.1 and 4.4.2) is presented, while the reader can find further examples showing the effectiveness of the other presented approaches in [P4] and [P8]. The following experimentation is carried out with the real-life mmW FR-2 environment shown in Section 5.1.1, with the aim of linearizing the active antenna array, while minimizing the complexity in the DPD learning path.

The provided linearization example is then presented in Fig. 5.9, for a very wide bandwidth of 800 MHz. It is noted that this signal bandwidth is not compliant with the latest 3GPP Release 15 [1], but is nonetheless chosen on purpose in order to push the performance boundaries of the DPD system. In the figure, the spectral curve without DPD has a TRP ACLR value of barely +21 dBc, which already demonstrates a very nonlinear operation point. Despite this, both ICM estimation methods

are able to generate a reliable ICM, which is then taken as an input for the SO learning algorithm. The SO learning algorithm provides also a reliable set of DPD coefficients, which are able to suppress the OoB nonlinear emissions to +31.5 dBc (autocorrelation estimation method) and +30 dBc (Bussgang's approximation method). Both of these values are also very close to the GN model, measured and presented as a reference technique. The average learning computational complexity of the proposed methods is, in turn, drastically reduced by up to 95%, when compared to the reference GN technique, as demonstrated in [P4] and [P8]. Additional convergence analyses of the proposed ICM methods can be found in [P4].

5.2 RF Verification in Digital Self-interference Cancellation

This second section deals with the verification of the proposed DSICs. Following an equivalent structure as taken before, the cascaded models are first tested, followed by the linear-in-parameters solutions.

5.2.1 RF Verification of Spline-based Cascaded Models

The various proposed cascaded models are tested and verified in the context of digital SI cancellation for IBFD devices through two different RF measurement environments. Firstly, the measured data originally used in [143] was utilized to test the SPH, SPW, and SPWH models. This data was extracted from a complete FR-1 IBFD prototype. The second measurement environment considers instead an FR-2 IBFD system, with an additional LNA in the RX chain to further demonstrate the effectiveness of the SPHW model. These experiments are explained within the following lines.

Digital SI Cancellation in a Complete FR-1 IBFD Prototype

The first measurement setup is conceptually depicted in Fig 5.10, containing a back-to-back relay antenna to provide as much TX-RX isolation as possible in the analog domain [117], [125]. In this example, the RF cancellation stage is omitted altogether, in contrast to other generic IBFD architectures. Omitting the RF cancellation stage is, however, done deliberately, as this results in a higher induced SI

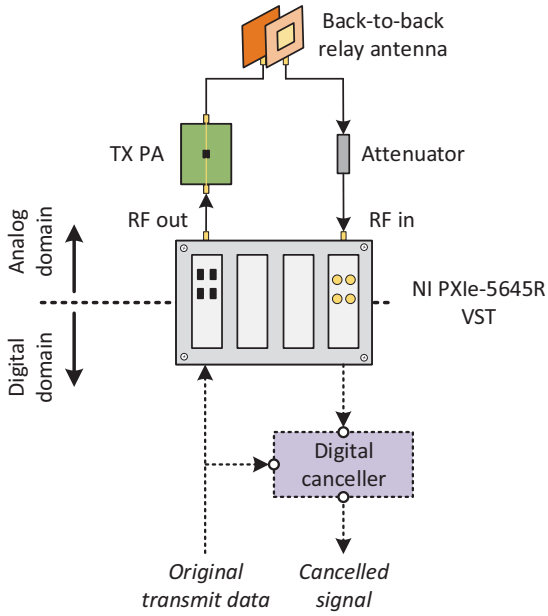
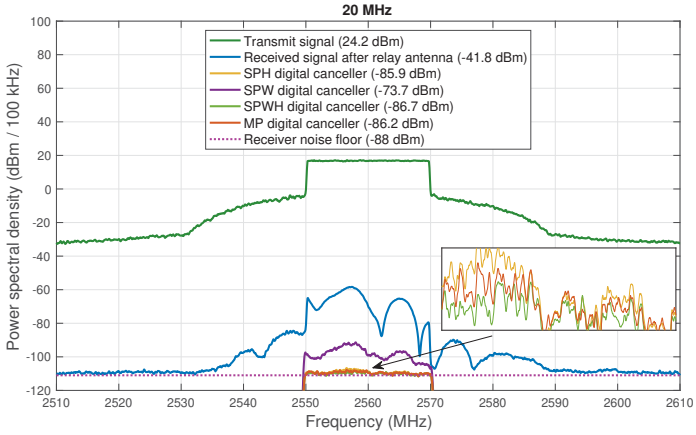


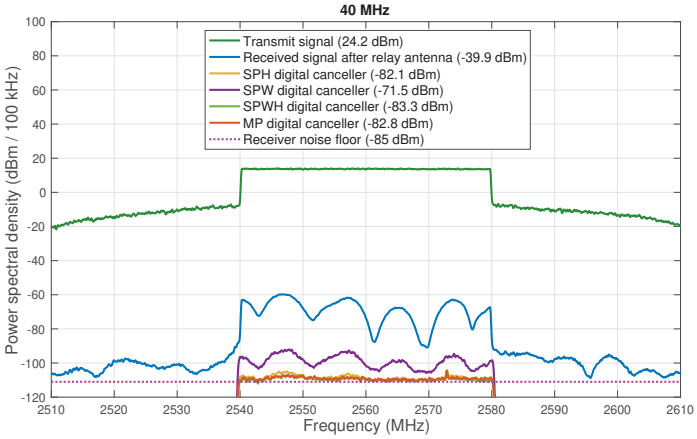
Figure 5.10 Conceptual structure of the FR-1 IBFD prototype utilized to evaluate the digital SI cancellation obtained with the spline-based cascaded structures.

interference before the ADC. As a consequence, a significant amount of SI suppression is still required, thus making this data set optimal to test the proposed SPH, SPW, and SPWH cascaded models, as reviewed in [P3, P5].

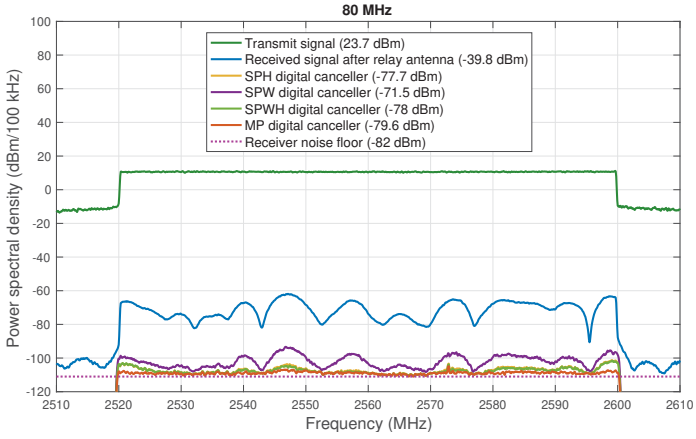
Three different SI cancellation examples are then presented in Fig 5.11, for the input signal bandwidths of 20, 40, and 80 MHz, respectively. The example power spectral densities (PSDs) show the cancellation performance of the SPH, SPW, and SPWH techniques, alongside a classical LMS-based MP model which is measured as a reference technique. In the three PSDs examples, the power of the RX signal after the back-to-back antenna isolation still ranges between -41.8 dBm (20 MHz) and -39.8 dBm (80 MHz), which is obviously above the RX noise floor level of -88 dBm, hence further digital cancellation is required. In the first bandwidth example of 20 MHz, the SPH and SPWH models are able to suppress the SI signal to around -86 dBm, which is only +2 dB higher than the RX noise floor. This result constitutes a digital SI cancellation of some +44 dB, which is at the same time comparable to also the +44 dB cancellation obtained with the MP model, regardless of the much lower complexity involved in the cascaded approaches. The cancellation of the SPW



(a)



(b)



(c)

Figure 5.11 PSDs after the digital SI cancellation stage in the IBFD prototype, when the cascaded SPH, SPW, and SPWH models are applied, for (a) 20 MHz, (b) 40 MHz, and (c) 80 MHz signal bandwidths.

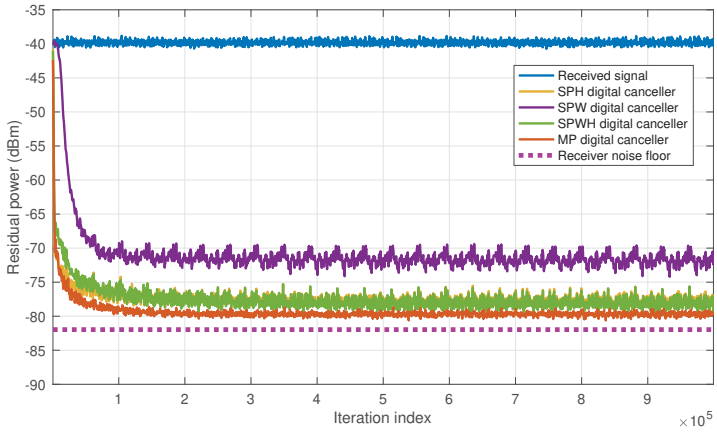


Figure 5.12 Residual error power with respect to the iteration index after applying the different DSIC algorithms, in the case of 80 MHz instantaneous bandwidth.

model, +32 dB, lies somewhat behind, but this result is expected since the structure of the considered IBFD prototype follows a Hammerstein system, in which the cancellation performance is non-optimal when modeled with a Wiener-type approach. The proposed cascaded approaches perform well also with the increased signal bandwidths of 40 and 80 MHz, as demonstrated within the figure. In the case of 40 MHz, the SPWH and MP models achieve an SI cancellation of some +43 dB, with the SPWH model even outperforming the MP approach. This result is followed by the SPH model, with +42 dB cancellation, and finally the SPW structure, with +32 dB cancellation. Similar conclusions apply for the latter case of 80 MHz, with an SI cancellation of +39 dB (SPWH and MP), +38 dB (SPH), and finally +32 dB (SPW).

A final convergence analysis of the proposed cascaded models is presented in Fig. 5.12, which shows the residual error power after cancellation as a function of the iteration index in the latter case when the bandwidth of 80 MHz is selected, while the former 20 MHz and 40 MHz cases can be found in [P3,P5]. It can be seen from the figure that all models essentially achieve a similar convergence speed, despite the cascaded models not using any type of orthogonalization. The figure also confirms the absolute levels of digital cancellation achieved by the different models, with the Wiener model lagging slightly behind, and with the SPH and SPWH models almost reaching the same performance as the reference MP model.

All in all, the cascaded approaches, especially SPH and SPWH, are able to successfully suppress the SI signal in the IBFD prototype, for all bandwidth cases of

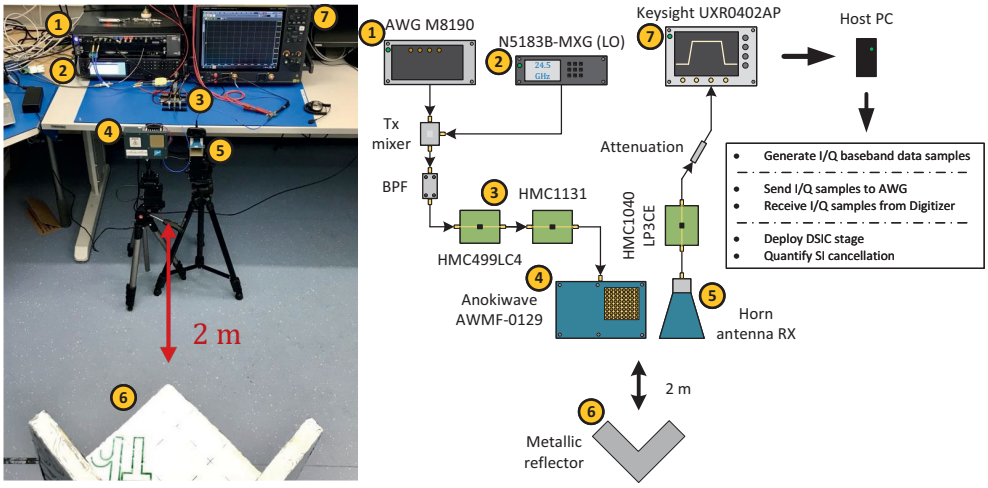


Figure 5.13 5G NR FR-2 measurement environment utilized to evaluate the digital SI cancellation obtained by the SPHW cascaded structure.

20-80 MHz, with cancellation results comparable to the reference LMS-based MP solution, regardless of the great complexity reductions of 82% (SPH and SPW), and 80% (SPWH) in the cancellation stage. Further details and measurements of this experiment can be found in [P3, P5].

Digital SI Cancellation in a Complete FR-2 Prototype

Second, the cascaded SPHW is tested within an FR-2 measurement system. A special IBFD setup incorporating a second LNA after the TX/RX antennas is incorporated to demonstrate the effectiveness of this model when modeling such structures. The considered RF measurement environment is depicted in Fig. 5.13. Specifically, the adopted mmW FR-2 measurement environment incorporates a 64-element active antenna array acting as the main TX PA, and an additional nonlinear AD HMC1040LP3CE LNA placed before the actual RX ADC. Moreover, a metallic reflector is situated at 2 meters from the TX/RX antennas, such that it produces a realistic reflection into the RX antenna. This reflection constitutes the main SI signal the IBFD device aims at suppressing. Since the power of the undesired signal reflection is not drastically high, no RF cancellation stage is considered, thus DSIC is required in the IBFD device for its successful operation. It is noted that there is no meaningful

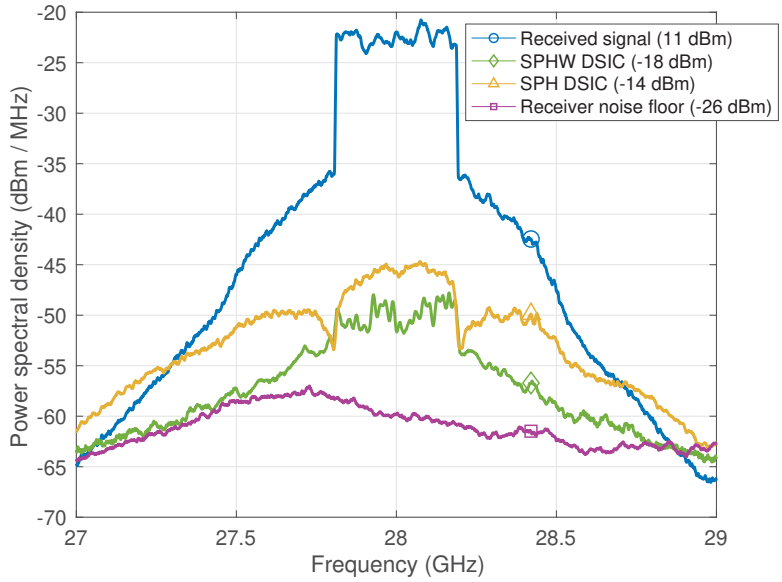


Figure 5.14 PSDs after the digital SI cancellation stage in the FR-2 measurement environment, when the cascaded SPHW model is applied, considering a signal bandwidth of 400 MHz. The SPH model is also presented for reference.

leakage from the array to the RX antenna, due to its highly directive beam-pattern. Corresponding signal upconversion/downconversion to/from 28 GHz and further signal conditioning stages are configured as explained in Section 5.1.1.

The SI cancellation results are then presented in Fig 5.14, showing also the performance of the SPH model for reference. In this example, the power of the interference signal coming from the metallic reflector is +11 dBm, which creates significant nonlinear distortions in the LNA, and thus additional DSIC processing is clearly needed to suppress this effect. After applying the digital cancellers, the SI signal is suppressed to -14 dBm with the SPH model, and to -18 dBm with the SPHW model. These results translate into a SI cancellation of +25 dB in the case of SPH, and up to +29 dB in the case of SPHW. With these type of structures, the performance enhancement achieved by the SPHW cascaded approach is clear, specifically providing +4 dB more of digital SI cancellation, since it is also capable of modeling the nonlinear response of the RX LNA. Further experimentation and testing of the SPHW model, including its convergence analysis, can be found in [P7].

5.2.2 RF Verification of Spline-based MP-LUT Models

In this last section, the cancellation performance of the linear-in-parameters SPMP model is tested and analyzed. The chosen RF measurement environment is the exact same as the one presented in Fig 5.10, where the output data is obtained from that same complete FR-1 IBFD prototype. For further details of this setup, refer to Section 5.2.1 and [P3, P5]. In this case, the SPMP model cancellation performance is tested with the widest signal bandwidth of 80 MHz, and compared to the classical MP and SPWH models for reference. The chosen amount of memory is the same as in the MP and SPWH models, and the LUT sizes are also the same as in the SPWH model.

The SI cancellation example result is then presented in Fig 5.15. Digital cancellation is indeed needed, since the power of the RX signal after the back-to-back antenna is still some -40 dBm. After applying the SPMP DSIC, the power of the SI signal is reduced to -78.2 dBm, which translates to +38.4 dB of digital SI cancellation. This result perfectly compares to the +39.9 dB of cancellation obtained with the twin MP model, regardless of the much lower complexity of the SPMP model. This result is also comparable to the +38.2 dB of cancellation obtained with the SPWH model. The convergence speed of the SPMP model is also similar to that of the SPWH and classical MP models, as can be verified from [P3,P7]. In terms of computational complexity in the cancellation stage, the SPMP model allows for a reduction of +42%, when compared to the classical MP model, while a more complete complexity analysis can be found in [P3]. Overall, these results indicate that high levels of efficient digital cancellation can indeed be achieved by the proposed SPMP technique.

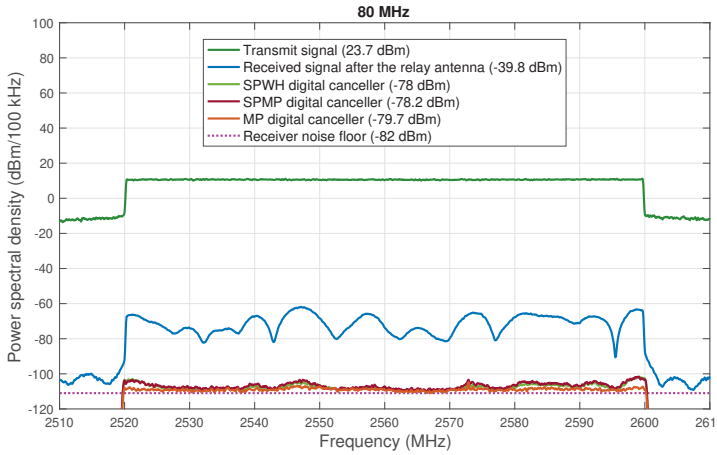


Figure 5.15 PSDs after the digital SI cancellation stage in the FR-1 measurement environment, when the SPMP DSIC is applied, considering a signal bandwidth of 80 MHz. The classical MP and proposed SPWH models are presented for reference.

6 CONCLUSION

THIS last chapter presents a summary of the main findings and results obtained within the research work carried out for this thesis, alongside with potential future directions of the studied topics. With 5G NR already well ramping up in modern mobile networks, and 6G being just around the corner, there are doubtlessly many possible directions for the future of DPD and IBFD. As reported within this work, the ever-increasing signal bandwidths, TX power efficiency, sampling rates, frequency ranges, etc. in radio transceivers call for low-complexity impairment compensation algorithms. Hence, there are still many relevant topics to be still explored, regardless of the several findings already presented here.

6.1 Main Findings

The principal aim of this thesis work was to develop low-complexity DPD and DSIC techniques applicable for modern wireless communications systems. On the one hand, the proposed DPD models targeted a power efficiency increase in the considered transmitter, while keeping the unwanted in-band and OoB emissions within specified levels. On the other hand, the proposed DSICs aimed at improving the spectral efficiency of the transmission by precisely modeling the unwanted SI signal which appears in any IBFD transceiver, and by removing its effect from the overall received signal in the cancellation stage. Since the transmit PA typically presents a highly saturated response, both of these approaches required a certain type of nonlinear modeling, and thus the modeling algorithms which were presented in this thesis could be used in both applications. To this end, the behavior of the nonlinear distortion injected by the transmit PA was analyzed, and several nonlinear models whose aim was to precisely estimate the PA nonlinear distortion were derived. Moreover, all of these models were supported by their corresponding mathematical derivations, and were tested and verified in all cases with commercial off-the-shelf

PAs, and through comprehensive real-life RF measurement environments, including state-of-the-art examples at FR-1 and FR-2.

The main findings and results found within this thesis work can be summarized as follows. First, Chapter 3 and publications [P1, P3, P5, P7] presented four different configurations of spline-interpolated cascaded structures to model the nonlinear behavior of a transmitter system, alongside their corresponding gradient-descent adaptive parameter learning solutions. Moreover, the cascaded approaches were configured with an injection-based scheme, which allowed to avoid gain ambiguities in the structures and also reduce the dynamic range of the spline control points. A vast amount of different experimental tests was provided, covering successful linearization and SI suppression in different scenarios. These results, together with the numerous complexity analyses provided, showed that the proposed cascaded models possess very appealing performance-complexity trade-offs. Specifically, the SPWH model usually behaved in a similar manner as other state-of-the-art reference solutions, such as the MP model, while drastically reducing the complexity involved by 50%. The performance of the SPH and SPW approaches typically followed behind, while achieving even a more drastic reduction in terms of computational complexity, in the order of 60%.

Next, in Chapter 4 and publications [P1, P2, P4, P6, P8], an efficient linear-in-parameters behavioral model was proposed, together with two alternative parameter learning solutions. The proposed model was implemented with the same spline interpolation processing scheme of Chapter 3, but now configured within a MP-type parallel branched structure to provide additional memory correction. In such a model, the parameter estimation was done with either a gradient descent-based adaptive learning solution, or with either the closed-loop-based forms of GN, SO, and BLMS. Additionally, as detailed in [P2, P6], the latter learning forms were also combined with efficient sign-based algorithms which signed different magnitudes in the learning updates to further reduce its computational complexity, while keeping an acceptable modeling performance. Specifically, the proposed signed models were shown to achieve a very close linearization behavior to that of the original unsigned version, while achieving complexity reductions in the order of 50-90%. Furthermore, to provide an even more drastic complexity reduction, several ICM methods were also proposed [P4, P8], in combination with the SO learning equation. In this case, the inherent low-complexity of the SO method, combined with the proposed

solutions, led to remarkably reduced complexities, which were in the order of 90%, while still maintaining a very close linearization performance to other reference approaches.

The performance of all aforementioned methods and models was tested and verified in Chapter 5. In the context of DPD, two state-of-the-art measurement environments were proposed, targeting FR-1 and FR-2 scenarios. The performance of the proposed algorithms was analyzed, and was also compared to other state-of-the-art solutions to demonstrate its effectiveness. In the experiments, impressive channel bandwidths of up to 800 MHz were reported. In all cases, the latest specified 3GPP Release 15 requirements of EVM and ACLR were fulfilled, regardless of the much lower complexity involved in the proposed systems. In the context of IBFD, the proposed models were configured as digital SI cancellers, and tested also in both FR-1 and FR-2 measurement environments. In all cases, the proposed DSIC solutions achieved high levels of digital SI suppression, with up to +45 dB and +32 dB when utilizing the FR-1 and FR-2 setups, respectively. These results are comparable in all cases to other state-of-the-art reference solutions, while the overall computational complexity in the modeling and learning paths was drastically reduced by more than 80%.

Altogether, this thesis showed that efficient PA linearization and digital SI cancellation can be achieved through the proposed techniques. The proposed digital methods are able to achieve levels of performance which are very close to, and even outperform in some cases, classical state-of-the-art solutions, regardless of their lower level of computational complexity involved. This was demonstrated through careful RF testing and experimentation, which complied in all cases with the latest international communications standards. In summary, this thesis, together with publications [P1-P8], provided several methods and solutions which were able to reach more efficient DPD systems, and also bring IBFD technology one step closer to commercially feasible implementations.

6.2 Future Work

This thesis work studied DPD and IBFD technologies, targeting the reduction of the computational complexity involved. In this context, this work unveiled many paths that still remain to be explored. Future directions and further research is hereby

summarized.

In the context of DPD, it is clearly a new trend to develop less complex predistorters which target a medium PA linearization performance, in contrast to classical methods which aimed at achieving high levels of distortion suppression. Adopting such low-complexity predistorters is particularly useful in mmW over-the-air arrays, where the nonlinear distortion is beam-dependent, and thus the DPD coefficients need to be estimated in the learning stage as quickly as possible as the beam is steered. In the existing literature, a lesser focus has been given to the learning complexity, thus it constitutes another area that requires further research.

Another whole topic that is somewhat unexplored is sampling reduction methods for DPD applications. Reducing the sampling rate can be achieved by two main approaches. The first method is band-limiting, which is already studied in DPD schemes. However, the overall complexity reduction of the band-limiting functions needs to be carefully assessed, as the upsampling and downsampling functions also add complexity to the system on their own. The second method is antialiasing, which has been studied within audio engineering applications, but its feasibility in DPD processing has so far yet to be studied. Such techniques operate instead at a lower sample rate all the time, thus avoiding the costly resampling processes needed in bandlimiting applications. All of these questions have not been addressed in the current literature, and would require further exploration.

In the context of IBFD, learning the associated parameters with the SO learning rule in combination with the ICM estimation methods reproduced in [P4, P8] is a natural continuation, which could further reduce the learning computational complexity in polynomial-type models. Furthermore, little attention has been given to the combination of IBFD with DPD. DPD can largely reduce the nonlinear distortion injected by the TX PA, thus simpler digital SI cancellation solutions could be adopted, which would reduce the modeling and learning associated complexities.

In addition, exploring IBFD technology in the context of mmW frequencies and active antenna arrays constitute a very interesting future research direction, with a preliminary idea already given in [P7]. It is well-known that very wide channel bandwidths, more difficult impedance matching, calibration challenges, higher frequency selectivity, etc. already appear in mmW frequencies, compared to other lower frequencies, and this will surely pose a challenge for IBFD and SI cancellation. Additionally, multi-antenna arrays inherently have nonlinear characteristics which are

beam-dependent, stemming from the load modulation phenomena of the PAs. The digital SI cancellers must take this into account, adapting their parameters as the beam of the array is steered. This is also clearly a future direction for IBFD technology that needs further research.

All in all, and despite the incremental progress that this thesis offered, there are still several key points that remain to be explored in the context of both DPD and IBFD. On the one hand, developing more efficient yet reliable DPD systems will allow for the minimization of cost and size of the utilized circuitry within the radio transceivers. On the other hand, it is crucial to keep researching and developing IBFD technology, with the ultimate aim of reaching full feasibility for its integration with commercial devices.

BIBLIOGRAPHY

[1] 3GPP Tech. Spec. 38.141-2, *NR; base station (BS) conformance testing, Part 2*, Mar. 2019.

[2] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital predistortion for hybrid MIMO transmitters," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 3, pp. 445–454, Jun. 2018, ISSN: 1932-4553. DOI: 10.1109/JSTSP.2018.2824981.

[3] M. Abdelaziz, "Reduced-complexity digital predistortion in flexible radio spectrum access," Ph.D. dissertation, Tampere University of Technology, 2017.

[4] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 1, pp. 280–293, 2018.

[5] M. Abdelaziz, L. Anttila, C. Tarver, K. Li, J. R. Cavallaro, and M. Valkama, "Low-complexity subband digital predistortion for spurious emission suppression in noncontiguous spectrum access," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 11, pp. 3501–3517, 2016.

[6] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive MIMO," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2017, pp. 6478–6482.

[7] M. Abi Hussein, V. A. Bohara, and O. Venard, "On the system level convergence of ILA and DLA for digital predistortion," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, 2012, pp. 870–874. DOI: 10.1109/ISWCS.2012.6328492.

[8] M. Abi Hussein, O. Venard, B. Feuvrie, and Y. Wang, "Digital predistortion for RF power amplifiers: State of the art and advanced approaches," in *2013 IEEE 11th International New Circuits and Systems Conference (NEWCAS)*, IEEE, 2013, pp. 1–4.

[9] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 4, pp. 1166–1174, Apr. 2012, ISSN: 0018-9480.

[10] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[11] S. Ahmadi, *5G NR: Architecture, technology, implementation, and operation of 3GPP new radio standards*. Academic Press, 2019.

[12] E. Ahmed, A. M. Eltawil, and A. Sabharwal, "Self-interference cancellation with nonlinear distortion suppression for full-duplex systems," in *2013 Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 1199–1203. DOI: 10.1109/ACSSC.2013.6810483.

[13] S. Amin, P. N. Landin, P. Händel, and D. Rönnow, "Behavioral modeling and linearization of crosstalk and memory effects in RF MIMO transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 810–823, 2014. DOI: 10.1109/TMTT.2014.2309932.

[14] M. S. Amjad and F. Dressler, "Experimental insights on software-based real-time SI cancellation for in-band full duplex FD relays," in *ICC 2019, 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6. DOI: 10.1109/ICC.2019.8762085.

[15] M. S. Amjad, H. Nawaz, K. Özsoy, Ö. Gürbüz, and I. Tekin, "A low-complexity full-duplex radio implementation with a single antenna," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2206–2218, 2018. DOI: 10.1109/TVT.2017.2765823.

[16] M. S. Amjad and O. Gurbuz, "Linear digital cancellation with reduced computational complexity for full-duplex radios," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2017, pp. 1–6.

[17] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Communications Letters*, vol. 8, no. 6, pp. 1653–1656, 2019.

[18] L. Anttila, "Digital front-end signal processing with widely-linear signal models in radio devices," Ph.D. dissertation, Tampere University of Technology, 2011.

[19] L. Anttila, P. Handel, O. Myllari, and M. Valkama, "Recursive learning-based joint digital predistorter for power amplifier and I/Q modulator impairments," *International Journal of Microwave and Wireless Technologies*, vol. 2, no. 2, pp. 173–182, 2010.

[20] L. Anttila, D. Korpi, V. Syrjälä, and M. Valkama, "Cancellation of power amplifier induced nonlinear self-interference in full-duplex transceivers," in *Proceedings 47th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov. 2013, pp. 1193–1198.

[21] L. Anttila, V. Lampu, S. A. Hassani, P. Pascual Campo, D. Korpi, M. Turunen, S. Pollin, and M. Valkama, "Full-duplexing with SDR devices: Algorithms, FPGA implementation and real-time results," *IEEE Transactions on Wireless Communications*, 2020.

[22] E. Aschbacher, "Digital pre-distortion of microwave power amplifiers," Ph.D. dissertation, Citeseer, 2005.

[23] A. C. M. Austin, A. Balatsoukas-Stimming, and A. Burg, "Digital predistortion of power amplifier non-linearities for full-duplex transceivers," in *2016 IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2016, pp. 1–5. DOI: 10.1109/SPAWC.2016.7536811.

[24] A. Balatsoukas-Stimming, "Non-linear digital self-interference cancellation for in-band full-duplex radios using neural networks," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5. DOI: 10.1109/SPAWC.2018.8445987.

[25] A. Balatsoukas-Stimming, A. C. Austin, P. Belanovic, and A. Burg, "Baseband and RF hardware impairments in full-duplex wireless systems: Experimental characterisation and suppression," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, pp. 1–11, 2015.

[26] C. Baquero Barneto, T. Riihonen, M. Turunen, L. Anttila, M. Fleischer, K. Stadius, J. Ryynänen, and M. Valkama, "Full-duplex OFDM radar with LTE and 5G NR waveforms: Challenges, solutions, and measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 10, pp. 4042–4054, Oct. 2019.

[27] E. J. Barlow, "Doppler radar," *Proceedings of the IRE*, vol. 37, no. 4, pp. 340–355, 1949. DOI: 10.1109/JRPROC.1949.231638.

[28] J. Becerra, A. Pérez-Hérnandez, M. Madero-Ayora, and C. Crespo-Cadenas, "A reduced-complexity direct learning architecture for digital predistortion through iterative pseudoinverse calculation," *IEEE Microwave and Wireless Components Letters*, 2021.

[29] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE Journal on Selected Areas in Communications*, vol. 1, no. 1, pp. 57–62, Jan. 1983.

[30] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *ACM SIGCOMM Computer Communication Review*, ACM, vol. 43, 2013, pp. 375–386.

[31] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[32] D. W. Bliss and Y. Rong, "Full-duplex self-interference mitigation performance in nonlinear channels," in *2014 48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 1696–1700. DOI: 10.1109/ACSSC.2014.7094756.

[33] H. Boutayeb and M. Darouach, "Recursive identification method for MISO Wiener-Hammerstein model," *IEEE transactions on automatic control*, vol. 40, no. 2, pp. 287–291, 1995.

[34] R. N. Braithwaite, "Wide bandwidth adaptive digital predistortion of power amplifiers using reduced order memory correction," in *2008 IEEE MTT-S International Microwave Symposium Digest*, IEEE, 2008, pp. 1517–1520.

[35] ——, "A comparison of indirect learning and closed loop estimators used in digital predistortion of power amplifiers," in *2015 IEEE MTT-S International Microwave Symposium*, IEEE, 2015, pp. 1–4.

[36] ——, "Fixed point considerations for digital predistortion of a RF power amplifier using recursive least square (RLS) estimation," in *2019 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, IEEE, 2019, pp. 1–3.

[37] R. N. Braithwaite and F. Luo, "General principles and design overview of digital predistortion," in *Digital Processing for Front End in Wireless Communication and Broadcasting*, 6, Cambridge Univ. Press, 2011, pp. 143–191.

[38] R. N. Braithwaite, "Digital predistortion of an RF power amplifier using a reduced Volterra series model with a memory polynomial estimator," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 10, pp. 3613–3623, Oct. 2017, ISSN: 1557-9670. DOI: 10.1109/TMTT.2017.2729513.

[39] R. N. Braithwaite, "Closed-loop digital predistortion (DPD) using an observation path with limited bandwidth," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 726–736, 2015. DOI: 10.1109/TMTT.2014.2388230.

[40] ——, "Digital predistortion of a RF power amplifier using a remote closed loop estimator," in *2020 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, 2020, pp. 42–45. DOI: 10.1109/PAWR46754.2020.9036001.

[41] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEEE Proceedings F - Communications, Radar and Signal Processing*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[42] A. Brihuega, M. Abdelaziz, L. Anttila, M. Turunen, M. Allén, T. Eriksson, and M. Valkama, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Transactions on Microwave Theory and Techniques*, pp. 1–1, 2020.

[43] A. Brihuega, M. Abdelaziz, M. Turunen, T. Eriksson, L. Anttila, and M. Valkama, "Linearizing active antenna arrays: Digital predistortion method and measurements," in *2019 IEEE MTT-S International Microwave Conference on Hardware and Systems for 5G and Beyond (IMC-5G)*, 2019, pp. 1–3. DOI: 10.1109/IMC-5G47857.2019.9160379.

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

[45] J. K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE transactions on vehicular technology*, vol. 39, no. 4, pp. 374–382, 1990.

[46] ——, "Optimum table spacing in predistorting amplifier linearizers," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 5, pp. 1699–1705, 1999.

[47] C.-F. Cheang, P.-I. Mak, and R. P. Martins, "A hardware-efficient feedback polynomial topology for DPD linearization of power amplifiers: Theory and FPGA validation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 9, pp. 2889–2902, 2018. DOI: 10.1109/TCSI.2017.2788082.

[48] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, 2014.

[49] T. Chen, J. Zhou, M. B. Dastjerdi, J. Diakonikolas, H. Krishnaswamy, and G. Zussman, "Demo abstract: Full-duplex with a compact frequency domain equalization-based RF canceller," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2017, pp. 972–973. DOI: 10.1109/INFCOMW.2017.8116516.

[50] M. Cheong, S. Werner, M. Bruno, J. Figueroa, J. Cousseau, and R. Wichman, "Adaptive piecewise linear predistorters for nonlinear power amplifiers with memory," *IEEE Transactions on Circuits and Systems I: Regular papers*, vol. 59, no. 7, pp. 1519–1532, 2012.

[51] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, ACM, 2010, pp. 1–12.

[52] S. Choi and E.-R. Jeong, "Digital predistortion based on combined feedback in MIMO transmitters," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1572–1575, 2012. DOI: 10.1109/LCOMM.2012.080312.120224.

[53] M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C. b. Chae, "Prototyping real-time full duplex radios," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 56–63, Sep. 2015, ISSN: 0163-6804. DOI: 10.1109/MCOM.2015.7263346.

[54] W. Chung, D. Hong, R. Wichman, and T. Riihonen, "Interference cancellation architecture for full-duplex system with GFDM signaling," in *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, 2016, pp. 788–792.

[55] G. Collins and D. W. Runton, "Nonlinear analysis of power amplifiers," *Microwave Journal*, vol. 50, no. 9, p. 164, 2007.

[56] I. Cosme, I. Fernandes, J. de Carvalho, and S. de Souza, "Block recursive matrix inverse," *arXiv preprint arXiv:1612.00001*, Nov. 2016.

[57] P. Crama and J. Schoukens, "Hammerstein–Wiener system estimator initialization," *Automatica*, vol. 40, no. 9, pp. 1543–1550, 2004.

[58] S. C. Cripps, *RF power amplifiers for wireless communications*. Artech house Norwood, MA, 2006, vol. 2.

[59] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, and Y. Selén, "5G radio access," *Ericsson review*, vol. 6, no. 1, 2014.

[60] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-advanced for mobile broadband*. Academic press, 2013.

[61] ——, *5G NR: The next generation wireless access technology*. Academic Press, 2020.

[62] A. I. Dalbah, O. Hammi, and A. Zerguine, "Hybrid look-up-tables based behavioral model for dynamic nonlinear power amplifiers," *IEEE Access*, vol. 8, pp. 53 240–53 249, 2020.

[63] A. I. Dalbah and O. Hammi, "Enhanced training technique for nested look-up table based behavioral modeling of nonlinear power amplifiers," in *2019 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2019, pp. 1–5. DOI: 10.1109/ICCSPA.2019.8713727.

[64] C. De Boor, *A practical guide to splines*. Springer, New York, 1978.

[65] D. De Caro, N. Petra, and A. G. Strollo, "Reducing lookup-table size in direct digital frequency synthesizers using optimized multipartite table method," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, no. 7, pp. 2116–2127, 2008.

[66] Ö. T. Demir and E. Björnson, "The Bussgang decomposition of nonlinear systems: Basic theory and MIMO extensions," *arXiv preprint arXiv:2005.01597*, 2020.

[67] G. Destino, O. Kursu, S. Tammelin, J. Haukipuro, M. Sonkki, T. Rahkonen, A. Pärssinen, M. Latva-aho, A. Korvala, and M. Pettissalo, "System analysis and design of mmw mobile backhaul transceiver at 28 GHz," in *2017 European Conference on Networks and Communications (EuCNC)*, IEEE, 2017, pp. 1–5.

[68] S. Dikmese, L. Anttila, P. P. Campo, M. Valkama, and M. Renfors, "Behavioral modeling of power amplifiers with modern machine learning techniques," in *2019 IEEE MTT-S International Microwave Conference on Hardware and Systems for 5G and Beyond (IMC-5G)*, IEEE, 2019, pp. 1–3.

[69] R. Dilli, "Analysis of 5G wireless systems in FR1 and FR2 frequency bands," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, IEEE, 2020, pp. 767–772.

[70] L. Ding and G. Zhou, "Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization," *IEEE Transactions on Vehicular Technology*, vol. 53, no. 1, pp. 156–162, 2004. DOI: 10.1109/TVT.2003.822001.

[71] L. Ding, "Digital predistortion of power amplifiers for wireless applications," Ph.D. dissertation, Georgia Institute of Technology, 2004.

[72] L. Ding, F. Mujica, and Z. Yang, "Digital predistortion using direct learning with reduced bandwidth feedback," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013, pp. 1–3. DOI: 10.1109/MWSYM.2013.6697388.

[73] L. Ding, R. Raich, and G. Tong Zhou, "A Hammerstein predistortion linearization design based on the indirect learning architecture," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 2002, pp. 2689–2692.

[74] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "Memory polynomial predistorter based on the indirect learning architecture," in *Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE*, IEEE, vol. 1, 2002, pp. 967–971.

[75] ——, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Transactions on communications*, vol. 52, no. 1, pp. 159–165, 2004.

[76] P. S. R. Diniz *et al.*, "Adaptive filtering: Algorithms and practical implementation," *The international series in Engineering and Computer Science*, pp. 147–150, 2008.

[77] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012, ISSN: 1536-1276.

[78] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, K. K. Ramakrishnan, C. W. Rice, and N. K. Shankaranarayanan, "Design and characterization of a full-duplex multiantenna system for WiFi networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 3, pp. 1160–1177, 2014. DOI: 10.1109/TVT.2013.2284712.

[79] M. Emara, P. Rosson, K. Roth, and D. Dassonville, "A full duplex transceiver with reduced hardware complexity," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8254958.

[80] M. Emara, K. Roth, L. G. Baltar, M. Faerber, and J. Nossek, "Nonlinear digital self-interference cancellation with reduced complexity for full duplex systems," in *WSA 2017; 21th International ITG Workshop on Smart Antennas*, VDE, 2017, pp. 1–6.

[81] S. Enserink, M. P. Fitz, K. Goverdhanam, Changyi Gu, T. R. Halford, I. Hossain, G. Karawasy, and O. Y. Takeshita, "Joint analog and digital interference cancellation," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, 2014, pp. 1–3. DOI: 10.1109/MWSYM.2014.6848493.

[82] H. Enzinger, K. Freiberger, G. Kubin, and C. Vogel, "Baseband Volterra filters with even-order terms: Theoretical foundation and practical implications," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 220–224. DOI: 10.1109/ACSSC.2016.7869028.

[83] C. Eun and E. J. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE transactions on signal processing*, vol. 45, no. 1, pp. 223–227, 1997.

[84] E. Eweda, "Analysis and design of a signed regressor LMS algorithm for stationary and nonstationary adaptive filtering with correlated Gaussian data," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 11, pp. 1367–1374, Nov. 1990. DOI: 10.1109/31.62411.

[85] 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 Microwave Magazine*, vol. 20, no. 5, pp. 35–49, May 2019, ISSN: 1527-3342. DOI: 10.1109/MMM.2019.2898020.

[86] M. Faulkner and M. Johansson, "Adaptive linearization using predistortion-experimental results," *IEEE Transactions on Vehicular Technology*, vol. 43, no. 2, pp. 323–332, May 1994, ISSN: 0018-9545. DOI: 10.1109/25.293651.

[87] X. Feng, "Efficient baseband digital predistortion techniques for linearizing power amplifier by taking into account nonlinear memory effect," Ph.D. dissertation, Université de Nantes, 2015.

[88] A. Filip, "Linear approximations to sqrt(x2 + y2) having equiripple error characteristics," *IEEE Transactions on Audio and Electroacoustics*, vol. 21, no. 6, pp. 554–556, 1973.

[89] M. Frerking, *Digital signal processing in communications systems*. Springer, 1994.

[90] A. Georgakopoulos, A. Margaris, K. Tsagkaris, and P. Demestichas, "Resource sharing in 5G contexts: Achieving sustainability with energy and resource efficiency," *IEEE Vehicular Technology Magazine*, vol. 11, no. 1, pp. 40–49, 2016.

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

[92] P. Gilabert, G. Montoro, and E. Bertran, "On the Wiener and Hammerstein models for power amplifier predistortion," in *2005 Asia-Pacific Microwave Conference Proceedings*, vol. 2, 2005, pp. 4–8. DOI: 10.1109/APMC.2005.1606491.

[93] P. L. Gilabert, A. Cesari, G. Montoro, E. Bertran, and J.-M. Dilhac, "Multi-lookup table FPGA implementation of an adaptive digital predistorter for linearizing RF power amplifiers with memory effects," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 2, pp. 372–384, 2008.

[94] P. L. Gilabert, G. Montoro, T. Wang, M. N. Ruiz, and J. A. Garcia, "Comparison of model order reduction techniques for digital predistortion of power amplifiers," in *2016 46th European Microwave Conference (EuMC)*, IEEE, 2016, pp. 182–185.

[95] P. L. Gilabert, D. Vegas, Z. Ren, G. Montoro, J. R. Perez-Cisneros, M. N. Ruiz, X. Si, and J. A. Garcia, "Design and digital predistortion linearization of a wideband outphasing amplifier supporting 200 MHz bandwidth," in *2020 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, IEEE, 2020, pp. 46–49.

[96] P. L. Gilabert, G. Montoro, and E. Bertran, "FPGA implementation of a real-time NARMA-based digital adaptive predistorter," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 58, no. 7, pp. 402–406, 2011. DOI: 10.1109/TCSII.2011.2158256.

[97] P. L. Gilabert, D. D. Silveira, G. Montoro, M. E. Gadringer, and E. Bertran, "Heuristic algorithms for power amplifier behavioral modeling," *IEEE Microwave and Wireless Components Letters*, vol. 17, no. 10, pp. 715–717, 2007.

[98] S. Gökceli, P. P. Campo, T. Levanen, J. Yli-Kaakinen, M. Turunen, M. Allén, T. Riihonen, A. Palin, M. Renfors, and M. Valkama, "SDR prototype for clipped and fast-convolution filtered OFDM for 5G new radio uplink," *IEEE Access*, vol. 8, pp. 89 946–89 963, 2020.

[99] G. H. Golub and C. F. Van Loan, "Matrix computations," in. JHU Press, 2012, vol. 3, pp. 238–239.

[100] T. Gotthans, R. Marsalek, M. Pospisil, J. Blumenstein, and G. Baudoin, "Experimental verification and comparison of polynomials and LUTs predistortion techniques," in *2015 25th International Conference Radioelektronika (RADIOELEKTRONIKA)*, IEEE, 2015, pp. 309–312.

[101] W. Greblicki, "Nonparametric approach to Wiener system identification," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 6, pp. 538–545, 1997.

[102] F. H. Gregorio, G. J. González, J. Cousseau, T. Riihonen, and R. Wichman, "Predistortion for power amplifier linearization in full-duplex transceivers without extra RF chain," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6563–6567. DOI: 10.1109/ICASSP.2017.7953421.

[103] M. Grimm, M. Allén, J. Marttila, M. Valkama, and R. Thomä, "Joint mitigation of nonlinear RF and baseband distortions in wideband direct-conversion receivers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 1, pp. 166–182, 2014. DOI: 10.1109/TMTT.2013.2292603.

[104] G. Gripenberg, S.-O. Londen, and O. Staffans, *Volterra integral and functional equations*, 34. Cambridge University Press, 1990.

[105] L. Guan and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 4, pp. 866–872, Apr. 2010, ISSN: 1557-9670. DOI: 10.1109/TMTT.2010.2041588.

[106] ——, "Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers," in *2011 Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits*, Apr. 2011, pp. 1–4. DOI: 10.1109/INMMIC.2011.5773325.

[107] L. Guan, C. Yu, and A. Zhu, "Bandwidth-constrained least squares-based model extraction for band-limited digital predistortion of RF power amplifiers," in *2012 Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits*, IEEE, 2012, pp. 1–3.

[108] N. Guan, N. Wu, and H. Wang, "Digital predistortion of wideband power amplifier with single undersampling ADC," *IEEE Microwave and Wireless Components Letters*, vol. 27, no. 11, pp. 1016–1018, Nov. 2017. DOI: 10.1109/LMWC.2017.2750059.

[109] ——, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microwave and Wireless Components Letters*, vol. 28, no. 10, pp. 921–923, Oct. 2018. DOI: 10.1109/LMWC.2018.2860790.

[110] H. Guo, J. Xu, S. Zhu, and S. Wu, "Realtime software defined self-interference cancellation based on machine learning for in-band full duplex wireless communications," in *2018 International Conference on Computing, Networking and Communications (ICNC)*, 2018, pp. 779–783. DOI: 10.1109/ICCNC.2018.8390351.

[111] D. A. Hall, "Understanding benefits of MIMO technology," *Microwaves and RF*, 2009.

[112] O. Hammi and F. M. Ghannouchi, "Twin nonlinear two-box models for power amplifiers and transmitters exhibiting memory effects with application to digital predistortion," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 8, pp. 530–532, Aug. 2009, ISSN: 1531-1309. DOI: 10.1109/LMWC.2009.2024848.

[113] O. Hammi, S. Carichner, B. Vassilakis, and F. M. Ghannouchi, "Power amplifiers' model assessment and memory effects intensity quantification using memoryless post-compensation technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 12, pp. 3170–3179, 2008. DOI: 10.1109/TMTT.2008.2006809.

[114] O. Hammi, F. M. Ghannouchi, S. Boumaiza, and B. Vassilakis, "A data-based nested LUT model for RF power amplifiers exhibiting memory effects," *IEEE microwave and wireless components letters*, vol. 17, no. 10, pp. 712–714, 2007.

[115] E. Hänsler and G. Schmidt, "Acoustic echo and noise control: A practical approach," in. John Wiley & Sons, 2005, vol. 40, pp. 43–46.

[116] S. S. Haykin, "Adaptive filter theory," in. Pearson Education India, 2008, pp. 226–227.

[117] M. Heino, S. N. Venkatasubramanian, C. Icheln, and K. Haneda, "Design of wavetraps for isolation improvement in compact in-band full-duplex relay antennas," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 3, pp. 1061–1070, 2015.

[118] S. Hesami, S. R. Aghdam, C. Fager, T. Eriksson, R. Farrell, and J. Dooley, "Single digital predistortion technique for phased array linearization," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2019, pp. 1–5.

[119] R. V. Hogg, J. McKean, and A. T. Craig, *Introduction to mathematical statistics*. Pearson Education, 2005.

[120] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-advanced*. John Wiley & Sons, 2011.

[121] S. Hong, J. Brand, J. I. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 114–121, 2014.

[122] X. Hong, Yu Gong, and Sheng Chen, "A Wiener model for memory high power amplifiers using B-spline function approximation," in *2011 17th International Conference on Digital Signal Processing (DSP)*, Jul. 2011, pp. 1–5. DOI: 10.1109/ICDSP.2011.6004878.

[123] C. Hoymann, D. Astely, M. Stattin, G. Wikstrom, J.-F. Cheng, A. Hoglund, M. Frenne, R. Blasco, J. Huschke, and F. Gunnarsson, "LTE release 14 outlook," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 44–49, 2016.

[124] H. Huang, J. Xia, and S. Boumaiza, "Novel parallel-processing-based hardware implementation of baseband digital predistorters for linearizing wideband 5G transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 9, pp. 4066–4076, 2020. DOI: 10.1109/TMTT.2020.2993236.

[125] Y. Huang and K. Boyle, *Antennas: from theory to practice*. John Wiley & Sons, 2008.

[126] M. Ibnkahla, N. Bershad, J. Sombrin, and F. Castanie, "Neural network modeling and identification of nonlinear channels with memory: Algorithms, applications, and analytic models," *IEEE Transactions on Signal Processing*, vol. 46, no. 5, pp. 1208–1220, 1998.

[127] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," *IEEE transactions on microwave theory and techniques*, vol. 54, no. 1, pp. 348–359, 2006.

[128] M. Jain, J. I. Choi, T. Kim, D. Bharadia, S. Seth, K. Srinivasan, P. Levis, S. Katti, and P. Sinha, "Practical, real-time, full duplex wireless," in *Proceedings of the 17th annual international conference on Mobile computing and networking*, 2011, pp. 301–312.

[129] F. Jalili, F. F. Tafuri, O. K. Jensen, Y. Li, M. Shen, and G. F. Pedersen, "Linearization trade-offs in a 5G mmWave active phased array OTA setup," *IEEE Access*, vol. 8, pp. 110 669–110 677, 2020. DOI: 10.1109/ACCESS.2020.3002348.

[130] P. Jardin and G. Baudoin, "Filter lookup table method for power amplifier linearization," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 3, pp. 1076–1087, 2007.

[131] J. Joung, C. K. Ho, K. Adachi, and S. Sun, "A survey on power-amplifier-centric techniques for spectrum-and energy-efficient wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 315–333, 2014.

[132] H. Kang, Y. Cho, and D. Youn, "On compensating nonlinear distortions of an OFDM system using an efficient adaptive predistorter," *IEEE Transactions on Communications*, vol. 47, no. 4, pp. 522–526, Apr. 1999.

[133] A. Katz, J. Wood, and D. Chokola, "The evolution of PA linearization: From classic feedforward and feedback through analog and digital predistortion," *IEEE Microwave Magazine*, vol. 17, no. 2, pp. 32–40, 2016. DOI: 10.1109/MMM.2015.2498079.

[134] H. H. Kelejian, "Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables," *Journal of the American Statistical Association*, vol. 66, no. 334, pp. 373–374, 1971.

[135] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, no. 23, p. 1, 2001.

[136] J. Kim, Y. Y. Woo, J. Moon, and B. Kim, "A new wideband adaptive digital predistortion technique employing feedback linearization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 2, pp. 385–392, 2008.

[137] J. Kim, P. Roblin, D. Chaillot, and Z. Xie, "A generalized architecture for the frequency-selective digital predistortion linearization technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 1, pp. 596–605, 2013.

[138] K. E. Kolodziej, B. T. Perry, and J. S. Herd, "In-band full-duplex technology: Techniques and systems survey," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 7, pp. 3025–3041, 2019. DOI: 10.1109/TMTT.2019.2896561.

[139] K. Komatsu, Y. Miyaji, and H. Uehara, "Frequency-domain Hammerstein self-interference canceller for in-band full-duplex OFDM systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2017.

[140] D. Korpi, "Full-duplex wireless: Self-interference modeling, digital cancellation, and system studies," Ph.D. dissertation, Tampere University of Technology, Dec. 2017.

[141] D. Korpi, M. Aghababaee, M. Piilila, L. Anttila, and M. Valkama, "Advanced architectures for self-interference cancellation in full-duplex radios: Algorithms and measurements," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 1553–1557. DOI: 10.1109/ACSSC.2016.7869639.

[142] D. Korpi, Y. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile in-band full-duplex radio: Algorithms and RF measurements," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–7.

[143] D. Korpi, M. Heino, C. Icheln, K. Haneda, and M. Valkama, "Compact in-band full-duplex relays with beyond 100 dB self-interference suppression: Enabling techniques and field measurements," *IEEE Transactions on Antennas and Propagation*, vol. 65, pp. 960–965, 2 Feb. 2017.

[144] D. Korpi, T. Huusari, Y. Choi, L. Anttila, S. Talwar, and M. Valkama, "Digital self-interference cancellation under nonideal RF components: Advanced algorithms and measured performance," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 286–290. DOI: 10.1109/SPAWC.2015.7227045.

[145] D. Korpi, J. Tamminen, M. Turunen, T. Huusari, Y.-S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Full-duplex mobile device: Pushing the limits," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 80–87, Sep. 2016.

[146] D. Korpi, M. Valkama, L. Anttila, and P. Pascual, "Spline-based nonlinear digital cancellation," in *In-band Full-Duplex Wireless Systems Handbook*, K. E. Kolodziej, Ed., Artech House, 2021, ch. 3.

[147] D. Korpi, S. Venkatasubramanian, T. Riihonen, L. Anttila, S. Otewa, C. Icheln, K. Haneda, S. Tretyakov, M. Valkama, and R. Wichman, "Advanced self-interference cancellation and multiantenna techniques for full-duplex radios," in *2013 Asilomar Conference on Signals, Systems and Computers*, 2013, pp. 3–8. DOI: 10.1109/ACSSC.2013.6810217.

[148] D. Korpi, L. Anttila, V. Syrjälä, and M. Valkama, "Widely linear digital self-interference cancellation in direct-conversion full-duplex transceiver," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1674–1687, 2014.

[149] D. Korpi, L. Anttila, and M. Valkama, "Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk," *EURASIP journal on wireless Communications and Networking*, vol. 2017, no. 1, pp. 1–15, 2017.

[150] D. Korpi, T. Riihonen, V. Syrjälä, L. Anttila, M. Valkama, and R. Wichman, "Full-duplex transceiver system calculations: Analysis of ADC and linearity challenges," *IEEE Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3821–3836, 2014.

[151] O. Kursu, M. E. Leinonen, G. Destino, N. Tervo, M. Sonkki, T. Rahkonen, A. Pärssinen, S. Tammelin, M. Pettissalo, and A. Korvala, "Design and measurement of a 5G mmW mobile backhaul transceiver at 28 GHz," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, pp. 1–11, 2018.

[152] Y. Kurzo, A. Burg, and A. Balatsoukas-Stimming, "Design and implementation of a neural network aided self-interference cancellation scheme for full-duplex radios," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 2018, pp. 589–593. DOI: 10.1109/ACSSC.2018.8645295.

[153] A. Kwan, O. Hammi, M. Helaoui, and F. M. Ghannouchi, "High performance wideband digital predistortion platform for 3G+ applications with better than 55dBc over 40 MHz bandwidth," in *2010 IEEE MTT-S International Microwave Symposium*, IEEE, 2010, pp. 1082–1085.

[154] A. K. Kwan, M. Helaoui, S. Boumaiza, M. R. Smith, and F. M. Ghannouchi, "Wireless communications transmitter performance enhancement using advanced signal processing algorithms running in a hybrid DSP/FPGA platform," *Journal of Signal Processing Systems*, vol. 56, no. 2-3, pp. 187–198, 2009.

[155] S. Lahti, P. P. Campo, V. Lampu, L. Anttila, M. Valkama, and T. D. Hämäläinen, "Implementation of a nonlinear self-interference canceller using high-level synthesis," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2020, pp. 1–5.

[156] P. Landin, M. Isaksson, and P. Handel, "Comparison of evaluation criteria for power amplifier behavioral modeling," in *2008 IEEE MTT-S Interna-*

tional Microwave Symposium Digest, 2008, pp. 1441–1444. DOI: 10.1109/MWSYM.2008.4633050.

[157] P. N. Landin and D. Rönnow, "RF PA modeling considering odd-even and odd order polynomials," in *2015 IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, 2015, pp. 1–6. DOI: 10.1109/SCVT.2015.7374233.

[158] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, 2014.

[159] N. Lashkarian, J. Shi, and M. Forbes, "A direct learning adaptive scheme for power-amplifier linearization based on Wirtinger calculus," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 12, pp. 3496–3505, 2014.

[160] H. Li, J. Van Kerrebrouck, O. Caytan, H. Rogier, J. Bauwelinck, P. Demeester, and G. Torfs, "Self-interference cancellation enabling high throughput short-reach wireless full-duplex communication," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6475–6486, 2018. DOI: 10.1109/TWC.2018.2859985.

[161] K. Li, A. Ghazi, C. Tarver, J. Boutellier, M. Abdelaziz, L. Anttila, M. Juntti, M. Valkama, and J. R. Cavallaro, "Parallel digital predistortion design on mobile GPU and embedded multicore CPU for mobile transmitters," *Journal of Signal Processing Systems*, vol. 89, no. 3, pp. 417–430, 2017.

[162] S. Li and R. D. Murch, "An investigation into baseband techniques for single-channel full-duplex wireless communication systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 9, pp. 4794–4806, 2014.

[163] K. Liang, J. Chen, and Y. E. Chen, "A quadratic-interpolated LUT-based digital predistortion technique for cellular power amplifiers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 3, pp. 133–137, Mar. 2014, ISSN: 1558-3791. DOI: 10.1109/TCSII.2013.2296194.

[164] B. van Liempd, B. Hershberg, B. Debaillie, P. Wambacq, and J. Craninckx, "An electrical-balance duplexer for in-band full-duplex with<-85dBm in-band distortion at +10dBm TX-power," in *ESSCIRC Conference 2015-41st*

European Solid-State Circuits Conference (ESSCIRC), IEEE, 2015, pp. 176–179.

[165] S. Lim and C. Eun, "Predistorter design for a memory-less nonlinear high power amplifier using the pth-order inverse method for OFDM systems," in *2005 International Symposium on Intelligent Signal Processing and Communication Systems*, 2005, pp. 217–220. DOI: 10.1109/ISPACS.2005.1595385.

[166] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Augmented Hammerstein predistorter for linearization of broad-band wireless transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 4, pp. 1340–1349, 2006.

[167] ——, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 1025–1033, 2004.

[168] T. Liu, Y. Ye, S. Yin, H. Chen, G. Xu, Y. Lu, and Y. Chen, "Digital predistortion linearization with deep neural networks for 5G power amplifiers," in *2019 European Microwave Conference in Central Europe (EuMCE)*, 2019, pp. 216–219.

[169] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 12, pp. 5235–5248, 2019.

[170] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, "Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018, ISSN: 0018-9480. DOI: 10.1109/TMTT.2018.2830772.

[171] Y. Liu, X. Quan, S. Shao, and Y. Tang, "Digital predistortion architecture with reduced ADC dynamic range," *Electronics Letters*, vol. 52, no. 6, pp. 435–437, 2016. DOI: 10.1049/el.2015.4174.

[172] Y. Liu, W. Pan, S. Shao, and Y. Tang, "A new digital predistortion for wideband power amplifiers with constrained feedback bandwidth," *IEEE Microwave and Wireless Components Letters*, vol. 23, no. 12, pp. 683–685, 2013.

[173] Y.-J. Liu, B. Lu, T. Cao, B.-H. Zhou, J. Zhou, and Y.-N. Liu, "On the robustness of look-up table digital predistortion in the presence of loop delay error," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 10, pp. 2432–2442, 2012.

[174] Y.-J. Liu, J. Zhou, W. Chen, and B.-H. Zhou, "A robust augmented complexity-reduced generalized memory polynomial for wideband RF power amplifiers," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2389–2401, 2013.

[175] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 742–758, 2014.

[176] L. C. Ludeman, *Fundamentals of digital signal processing*. Harper & Row New York, 1986.

[177] F.-L. Luo, *Digital front-end in wireless communications and broadcasting: circuits and signal processing*. Cambridge University Press, 2011.

[178] Q. Luo, X. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 1, pp. 301–315, 2020.

[179] Y. Ma, Y. Yamao, Y. Akaiwa, and K. Ishibashi, "Wideband digital predistortion using spectral extrapolation of band-limited feedback signal," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 7, pp. 2088–2097, Jul. 2014, ISSN: 1549-8328. DOI: 10.1109/TCSI.2013.2295897.

[180] Y. Ma, Y. Yamao, Y. Akaiwa, and C. Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 11, pp. 3961–3973, 2013. DOI: 10.1109/TMTT.2013.2281962.

[181] D. P. Mandic and V. Su Lee Goh, *Complex valued nonlinear adaptive filters: noncircularity, widely linear and neural models*. Wiley, 2009.

[182] E. Manuzzato, J. Tamminen, M. Turunen, D. Korpi, F. Granelli, and M. Valkama, "Digitally-controlled electrical balance duplexer for transmitter receiver isolation in full-duplex radio," in *European Wireless 2016; 22th European Wireless Conference*, VDE, 2016, pp. 1–8.

[183] C. Mateo, P. L. Carro, P. Garcia-Ducar, J. de Mingo, and I. Salinas, "Digital predistortion based on B-spline CPWL models in a RoF LTE mobile fronthaul," in *2017 12th European Microwave Integrated Circuits Conference (EuMIC)*, 2017, pp. 396–399. DOI: 10.23919/EuMIC.2017.8230742.

[184] V. J. Mathews and G. Sicuranza, *Polynomial signal processing*. John Wiley & Sons, Inc., 2000.

[185] J. Minkoff, "The role of AM-to-PM conversion in memoryless nonlinear systems," *IEEE Transactions on Communications*, vol. 33, no. 2, pp. 139–144, 1985.

[186] F. Mkadem, M. B. Ayed, S. Boumaiza, J. Wood, and P. Aaen, "Behavioral modeling and digital predistortion of power amplifiers with memory using two hidden layers artificial neural networks," in *2010 IEEE MTT-S International Microwave Symposium*, 2010, pp. 656–659. DOI: 10.1109/MWSYM.2010.5517039.

[187] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for RF power amplifier behavioral modeling and digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 4, pp. 913–923, 2011. DOI: 10.1109/TMTT.2010.2098041.

[188] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-band complexity-reduced generalized-memory-polynomial power-amplifier digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 6, pp. 1763–1774, 2016.

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

[190] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, 2017.

[191] C. Mollen, E. G. Larsson, U. Gustavsson, T. Eriksson, and R. W. Heath, "Out-of-band radiation from large antenna arrays," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 196–203, Apr. 2018, ISSN: 0163-6804. DOI: 10.1109/MCOM.2018.1601063.

[192] G. Montoro, P. L. Gilabert, E. Bertran, A. Cesari, and J. A. Garcia, "An LMS-based adaptive predistorter for cancelling nonlinear memory effects in RF power amplifiers," in *2007 Asia-Pacific Microwave Conference*, IEEE, 2007, pp. 1–4.

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

[194] Y. Nagata, "Linear amplification technique for digital mobile communications," in *IEEE 39th Vehicular Technology Conference*, IEEE, 1989, pp. 159–164.

[195] M. D. Nair, R. Giofre, P. Colantonio, and F. Giannini, "NARMA based novel closed loop digital predistortion using Moore-Penrose inverse technique," in *2016 11th European Microwave Integrated Circuits Conference (EuMIC)*, IEEE, 2016, pp. 405–408.

[196] E. Ng, Y. Beltagy, G. Scarlato, A. B. Ayed, P. Mitran, and S. Boumaiza, "Digital predistortion of millimeter-wave RF beamforming arrays using low number of steering angle-dependent coefficient sets," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 11, pp. 4479–4492, 2019.

[197] M. O'Droma, S. Meza, and Y. Lei, "New modified Saleh models for memoryless nonlinear power amplifier behavioural modelling," *IEEE Communications Letters*, vol. 13, no. 6, pp. 399–401, 2009.

[198] S. J. Orfanidis, *Introduction to signal processing*. Prentice-Hall, Inc., 1995.

[199] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *2008 IEEE International Symposium on Wireless Communication Systems*, 2008, pp. 309–313. DOI: 10.1109/ISWCS.2008.4726067.

[200] P. Pascual Campo, V. Lampu, A. Meirhaeghe, J. Boutellier, L. Anttila, and M. Valkama, "Digital predistortion for 5G small cell: GPU implementation and RF measurements," *Journal of Signal Process. Systems*, pp. 1–12, 2019.

[201] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE transactions on microwave theory and techniques*, vol. 53, no. 4, pp. 1150–1163, 2005.

[202] E. Peralta, T. Levanen, T. Ihalainen, S. Nielsen, M. H. Ng, M. Renfors, and M. Valkama, "5G new radio base-station sensitivity and performance," in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, IEEE, 2018, pp. 1–6.

[203] K. Petersen and M. Pedersen, "The matrix cookbook," *Technical Univ. Denmark, Kongens Lyngby, Denmark, Tech. Rep*, vol. 3274, 2012.

[204] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[205] D. M. Pozar, *Microwave engineering*. John wiley & sons, 2011.

[206] H. Prautzsch, W. Boehm, and M. Paluszny, *Bézier and B-spline techniques*. Springer Science & Business Media, 2002.

[207] C. D. Presti, D. F. Kimball, and P. M. Asbeck, "Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA PA module," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3, pp. 604–618, 2012.

[208] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IRE Transactions on Information Theory*, vol. 4, no. 2, pp. 69–72, 1958.

[209] T. Rahkonen, S. Hietakangas, and J. Aikio, "AM-PM distortion caused by transistor's signal-dependent input impedance," in *2011 20th European Conference on Circuit Theory and Design (ECCTD)*, IEEE, 2011, pp. 833–836.

[210] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 4, pp. 1567–1592, Mar. 2013, ISSN: 1553-877X. DOI: 10.1109/SURV.2013.021313.00164.

[211] B. S. Rao and D. E. N. Davies, "Some preliminary results in a study of a twin-channel frequency-modulated radar," *IETE Journal of Research*, vol. 10, no. 11, pp. 516–523, 1964. DOI: 10.1080/03772063.1964.11485094.

[212] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.

[213] M. Rawat, F. M. Ghannouchi, and K. Rawat, "Three-layered biased memory polynomial for dynamic modeling and predistortion of transmitters with memory," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 3, pp. 768–777, 2012.

[214] I. Reed, "On a moment theorem for complex Gaussian processes," *IRE Transactions on Information Theory*, vol. 8, no. 3, pp. 194–195, 1962.

[215] O. Rioul and J. C. Magossi, "On Shannon's formula and Hartley's rule: Beyond the mathematical coincidence," *Entropy*, vol. 16, no. 9, pp. 4892–4910, 2014.

[216] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, 2014.

[217] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014, ISSN: 0733-8716.

[218] N. Safari, P. Fedorenko, J. S. Kenney, and T. Roste, "Spline-based model for digital predistortion of wide-band signals for high power amplifier linearization," in *2007 IEEE/MTT-S International Microwave Symposium*, 2007, pp. 1441–1444. DOI: 10.1109/MWSYM.2007.380504.

[219] M. Sakai, H. Lin, and K. Yamashita, "Adaptive cancellation of self-interference in full-duplex wireless with transmitter IQ imbalance," in *2014 IEEE Global Communications Conference*, 2014, pp. 3220–3224. DOI: 10.1109/GLOCOM.2014.7037302.

[220] G. Sanfilippo, O. Galinina, S. Andreev, S. Pizzi, and G. Araniti, "A concise review of 5G new radio capabilities for directional access at mmWave frequencies," in *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, Springer, 2018, pp. 340–354.

[221] A. Sano and L. Sun, "Identification of Hammerstein-Wiener system with application to compensation for nonlinear distortion," in *41st SICE Annual Conference*, 2002, pp. 1521–1526.

[222] M. G. Sarret, M. Fleischer, G. Berardinelli, N. H. Mahmood, P. Mogensen, and H. Heinz, "On the potential of full duplex performance in 5G ultra-dense small cell networks," in *2016 24th European Signal Processing Conference (EUSIPCO)*, IEEE, 2016, pp. 764–768.

[223] W. K. Saunders, "Post-war developments in continuous-wave and frequency-modulated radar," *IRE Transactions on Aerospace and Navigational Electronics*, vol. ANE-8, no. 1, pp. 7–19, 1961. DOI: 10.1109/TANE3.1961.4201771.

[224] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Nonlinear spline adaptive filtering," *Signal Processing*, vol. 93, no. 4, pp. 772–783, 2013.

[225] ——, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties," *Signal Processing*, vol. 100, pp. 112–123, 2014.

[226] ——, "Novel cascade spline architectures for the identification of nonlinear systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 7, pp. 1825–1835, 2015.

[227] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 5, pp. 285–291, 1976. DOI: 10.1109/TCS.1976.1084219.

[228] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals*. Cambridge university press, 2010.

[229] K.-H. Seah, C. Papavassiliou, G. Constantinides, *et al.*, "Digital polar transmitter for ultra-wideband system using OFDM modulation," *Electronics Letters*, vol. 43, no. 8, pp. 466–468, 2007.

[230] M. M. Shammasi and S. M. Safavi, "Performance of a predistorter based on saleh model for OFDM systems in HPA nonlinearity," in *2012 14th International Conference on Advanced Communication Technology (ICACT)*, IEEE, 2012, pp. 148–152.

[231] F. J. Soriano-Irigaray, J. S. Fernandez-Prat, F. J. Lopez-Martinez, E. Martos-Naya, O. Cobos-Morales, and J. T. Entrambasaguas, "Adaptive self-interference cancellation for full duplex radio: Analytical model and experimental validation," *IEEE Access*, vol. 6, pp. 65 018–65 026, 2018. DOI: 10.1109/ACCESS.2018.2878448.

[232] W.-H. Steeb and Y. Hardy, *Matrix calculus and Kronecker product: a practical approach to linear and multilinear algebra second edition*. World Scientific Publishing Company, 2011.

[233] A. G. Stove, "Linear FMCW radar techniques," in *IEE Proceedings F (Radar and Signal Processing)*, IET, vol. 139, 1992, pp. 343–350.

[234] G. Su, W. Chen, S. Zhang, and F. M. Ghannouchi, "A robust and low sampling rate digital predistortion algorithm for broadband PA modeling and predistortion," in *WAMICON 2014*, IEEE, 2014, pp. 1–4.

[235] M. A. Tafreshi, M. Koskela, D. Korpi, P. Jääskeläinen, M. Valkama, and J. Takala, "Software defined radio implementation of adaptive nonlinear digital self-interference cancellation for mobile inband full-duplex radio," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016, pp. 733–737. DOI: 10.1109/GlobalSIP.2016.7905939.

[236] K. Takahashi and S. Mori, "A new normalized signed regressor LMS algorithm," in *[Proceedings] Singapore ICCS/ISITA '92*, 1992, 1181–1185 vol.3. DOI: 10.1109/ICCS.1992.255075.

[237] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010, ISSN: 0018-9480. DOI: 10.1109/TMTT.2010.2047920.

[238] M. F. Triola, W. M. Goodman, R. Law, and G. Labute, *Elementary statistics*. Pearson Boston, MA, 2014, vol. 20151028.

[239] J. A. Tropp, "Recovery of short, complex linear combinations via l1 minimization," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1568–1570, Apr. 2005. DOI: 10.1109/TIT.2005.844057.

[240] C. F. Van Loan, "The ubiquitous Kronecker product," *Journal of Computational and Applied Mathematics*, vol. 123, no. 1-2, pp. 85–100, 2000.

[241] M. Vetterli, J. Kovacevic, and V. K. Goyal, *Foundations of signal processing*. Cambridge University Press, 2014.

[242] J. Vörös, "An iterative method for Hammerstein-Wiener systems parameter identification," *Journal of electrical engineering*, vol. 55, no. 11-12, pp. 328–331, 2004.

[243] J. H. Vuolevi, T. Rahkonen, and J. P. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Transactions on microwave theory and techniques*, vol. 49, no. 8, pp. 1383–1389, 2001.

[244] M. Z. Waheed, D. Korpi, L. Anttila, A. Kiayani, M. Kosunen, K. Stadius, P. P. Campo, M. Turunen, M. Allén, J. Ryynänen, *et al.*, "Passive intermodulation in simultaneous transmit-receive systems: Modeling and digital cancellation methods," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 9, pp. 3633–3652, 2020.

[245] M. Z. Waheed, P. Pascual Campo, D. Korpi, A. Kiayani, L. Anttila, and M. Valkama, "Digital cancellation of passive intermodulation in FDD

transceivers," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, IEEE, 2018, pp. 1375–1381.

[246] H. Wang, G. Li, C. Zhou, W. Tao, F. Liu, and A. Zhu, "1-bit observation for direct-learning-based digital predistortion of RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 7, pp. 2465–2475, Jul. 2017. DOI: 10.1109/TMTT.2016.2642945.

[247] S. Wang, M. Roger, and C. Lelandais-Perrault, "Impacts of crest factor reduction and digital predistortion on linearity and power efficiency of power amplifiers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 66, no. 3, pp. 407–411, 2018.

[248] T. Wang and J. Ilow, "Compensation of nonlinear distortions with memory effects in OFDM transmitters," in *IEEE Global Telecommunications Conference*, IEEE, 2004, pp. 2398–2403.

[249] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G massive MIMO wireless transmitters based on indirect identification of power amplifier behavior with OTA tests," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 1, pp. 316–328, 2020. DOI: 10.1109/TMTT.2019.2944828.

[250] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low feedback sampling rate digital predistortion for wideband wireless transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 11, pp. 3528–3539, Nov. 2016, ISSN: 0018-9480. DOI: 10.1109/TMTT.2016.2602216.

[251] 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 Transactions on Microwave Theory and Techniques*, vol. 65, no. 11, pp. 4274–4284, 2017.

[252] R. J. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Computation*, vol. 2, no. 4, pp. 490–501, 1990. DOI: 10.1162/neco.1990.2.4.490.

[253] Y. Y. Woo, J. Kim, J. Yi, S. Hong, I. Kim, J. Moon, and B. Kim, "Adaptive digital feedback predistortion technique for linearizing power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 5, pp. 932–940, 2007.

[254] D. Wu, C. Zhang, S. Gao, and D. Chen, "A digital self-interference cancellation method for practical full-duplex radio," in *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2014, pp. 74–79. DOI: 10.1109/ICSPCC.2014.6986155.

[255] X. Wu, N. Zheng, X. Yang, J. Shi, and H. Chen, "A spline-based Hammerstein predistortion for 3G power amplifiers with hard nonlinearities," in *2010 2nd International Conference on Future Computer and Communication*, vol. 3, May 2010. DOI: 10.1109/ICFCC.2010.5497423.

[256] X. Wu, J. Shi, and H. Chen, "On the numerical stability of RF power amplifier's digital predistortion," in *2009 15th Asia-Pacific Conference on Communications*, IEEE, 2009, pp. 430–433.

[257] Y.-Y. Xiao, "Stability analysis of closed-loop algorithms for cancelling nonlinear memory effects," *IEEE Transactions on Signal Processing*, vol. 65, pp. 1477–1485, 2017.

[258] S. Yao, H. Qian, K. Kang, and M. Shen, "A recursive least squares algorithm with reduced complexity for digital predistortion linearization," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, 2013, pp. 4736–4739.

[259] Q. You, L. Gu, and J. Liu, "Improved variable step-size LMS for digital predistortion in wideband power amplifiers," in *2018 8th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, IEEE, 2018, pp. 116–119.

[260] M. Younes, O. Hammi, A. Kwan, and F. M. Ghannouchi, "An accurate complexity-reduced "PLUME" model for behavioral modeling and digital predistortion of RF power amplifiers," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 4, pp. 1397–1405, Apr. 2011, ISSN: 0278-0046. DOI: 10.1109/TIE.2010.2049717.

[261] C. Yu, W. Cao, Y. Guo, and A. Zhu, "Digital compensation for transmitter leakage in non-contiguous carrier aggregation applications with FPGA implementation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 12, pp. 4306–4318, 2015.

[262] S. Zhang, W. Chen, and Z. Feng, "Low sampling rate digital predistortion of power amplifier assisted by bandpass RF filter," in *2012 Asia Pacific Microwave Conference Proceedings*, IEEE, 2012, pp. 962–964.

[263] X.-D. Zhang, *Matrix analysis and applications*. Cambridge University Press, 2017.

[264] H. Zhi-yong, G. Jian-hua, G. Shu-Jian, and W. Gang, "An improved look-up table predistortion technique for HPA with memory effects in OFDM systems," *IEEE Transactions on Broadcasting*, vol. 52, no. 1, pp. 87–91, 2006.

[265] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 120–133, 2007. DOI: 10.1109/TSP.2006.882058.

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

[267] ——, "An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers," in *2006 IEEE Annual Wireless and Microwave Technology Conference*, Dec. 2006, pp. 1–5. DOI: 10.1109/WAMICON.2006.351917.

[268] A. Zhu, M. Wren, and T. J. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers," in *IEEE MTT-S International Microwave Symposium Digest, 2003*, vol. 2, 2003, pp. 787–79. DOI: 10.1109/MWSYM.2003.1212488.

[269] A. Zhu, "Digital predistortion and its combination with crest factor reduction," in *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*, F.-L. Luo, Ed., Cambridge: Cambridge University Press, 2011, ch. 9, pp. 244–279.

[270] ——, "Decomposed vector rotation-based behavioral modeling for digital predistortion of RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 737–744, 2015. DOI: 10.1109/TMTT.2014.2387853.

[271] A. Zhu, J. C. Pedro, and T. R. Cunha, "Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 5, pp. 813–821, 2007.

[272] A. Zhu, J. C. Pedro, and T. C. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 12, pp. 4323–4332, 2006.

APPENDIX AM/AM RESPONSES OF THE TESTED PA SYSTEMS

THIS appendix provides representative measured AM/AM curves of all the PA systems utilized for the RF verification and validation of the proposed DPD and IBFD methods, which were described within this Dissertation and also within the attached papers.

Specifically, the AM/AM responses of the Mini-circuits ZHL-4240 PA unit, the Skyworks SKY66293-21 PA unit, and the Anokiwave AWMF-0129 active antenna array, are featured in Fig. 1, Fig. 2, and Fig. 3, respectively. Further measurement parameters can be found in corresponding figure captions.

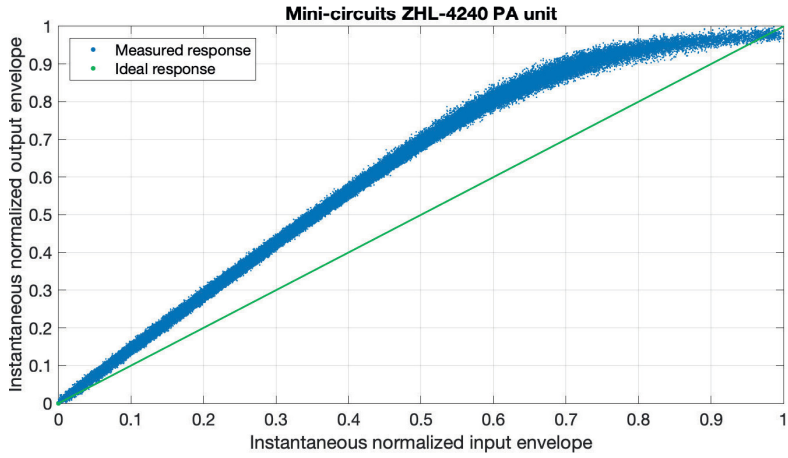


Figure 1 Instantaneous normalized AM/AM response of the Mini-circuits ZHL-4240 PA system, obtained with a 5G NR signal bandwidth of 100 MHz, an RF center frequency of 3.5 GHz, an output power of +27 dBm, and a PAPR of 8 dB (0.01% CCDF). This PA unit has a gain of +41 dB and a 1-dB compression point of +31 dBm.

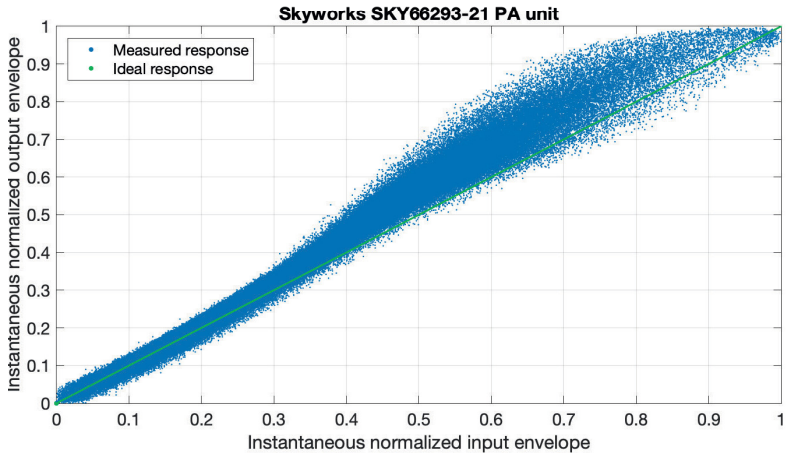


Figure 2 Instantaneous normalized AM/AM response of the Skyworks SKY66293-21 PA system, obtained with a 5G NR signal bandwidth of 100 MHz, an RF center frequency of 3.65 GHz, an output power of +27 dBm, and a PAPR of 8 dB (0.01% CCDF). This PA unit has a gain of +34 dB and a 1-dB compression point of +31.5 dBm.

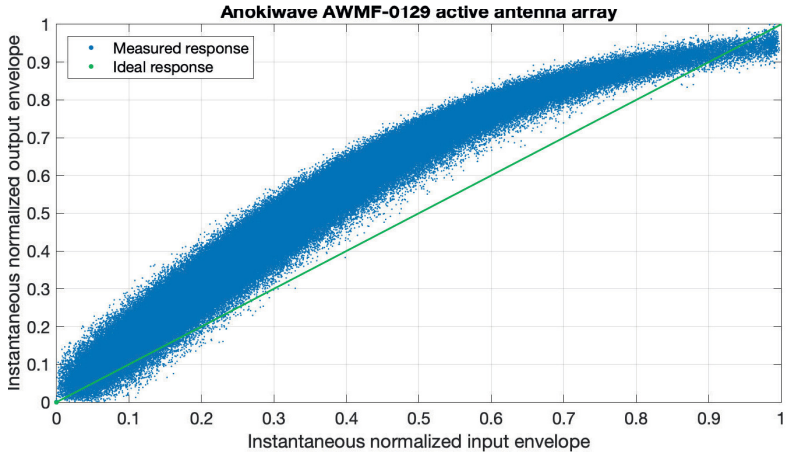


Figure 3 Instantaneous normalized AM/AM response of the Anokiwave AWMF-0129 active antenna array, obtained with a 5G NR signal bandwidth of 800 MHz, an RF center frequency of 28 GHz, an EIRP of +40.5 dBm, since an OTA measurement system is considered, and a PAPR of 8 dB (0.01% CCDF). This active antenna array has a 1-dB compression point of +41 dBm.

PUBLICATIONS

PUBLICATION 1

Gradient-adaptive spline-interpolated LUT methods for low-complexity digital predistortion

P. Pascual Campo, A. Brihuega, L. Anttila, M. Turunen, D. Korpi, M. Allén and M. Valkama

IEEE Transactions on Circuits and Systems I: Regular Papers 68.1 (2021), 336–349
DOI: 10.1109/TCSI.2020.3034825

Publication reprinted with the permission of the copyright holders

Gradient-Adaptive Spline-Interpolated LUT Methods for Low-Complexity Digital Predistortion

Pablo Pascual Campo, Alberto Brihuega, *Student Member, IEEE,* Lauri Anttila, *Member, IEEE,* Matias Turunen, Dani Korpi, *Member, IEEE,* Markus Allén, and Mikko Valkama, *Senior Member, IEEE*

***Abstract*—In this paper, new digital predistortion (DPD) solutions for power amplifier (PA) linearization are proposed, with particular emphasis on reduced processing complexity in future 5G and beyond wideband radio systems. The first proposed method, referred to as the spline-based Hammerstein (SPH) approach, builds on complex spline-interpolated lookup table (LUT) followed by a linear finite impulse response (FIR) filter. The second proposed method, the spline-based memory polynomial (SMP) approach, contains multiple parallel complex spline-interpolated LUTs together with an input delay line such that more versatile memory modeling can be achieved. For both structures, gradient-based learning algorithms are derived to efficiently estimate the LUT control points and other related DPD parameters. Large set of experimental results are provided, with specific focus on 5G New Radio (NR) systems, showing successful linearization of multiple PA samples as well as a 28 GHz active antenna array, incorporating channel bandwidths up to 200 MHz. Explicit performance-complexity comparisons are also reported between the SPH and SMP DPD systems and the widely-applied ordinary memory-polynomial (MP) DPD solution. The results show that the linearization capabilities of the proposed methods are very close to that of the ordinary MP DPD, particularly with the proposed SMP approach, while having substantially lower processing complexity.**

***Index Terms*—Digital predistortion, power amplifier, spline interpolation, Hammerstein, memory polynomial, lookup table, nonlinear distortion, behavioral modeling, EVM, ACLR**

I. Introduction

MODERN radio communication systems, such as the 4G LTE/LTE-Advanced and the emerging 5G New Radio (NR) mobile networks, build on multicarrier modulation, most notably orthogonal frequency division multiplexing (OFDM) [1]. OFDM waveforms are known to contain high peak-to-average power-ratio (PAPR) [2], [3], which complicates utilizing highly nonlinear power amplifiers (PAs) in transmitters operating close to saturation [2], [4], [5]. Digital predistortion (DPD) is, generally, a well-established approach to control the unwanted emissions and nonlinear distortion stemming from nonlinear PAs, see, e.g., [2], [4], [6]–[9] and references therein. Especially when combined with appropriate PAPR reduction methods [10], DPD based systems can largely improve the transmitter power efficiency while keeping the unwanted emissions within specified limits.

Some of the most common approaches in PA direct modeling as well as DPD processing are the memory polynomial (MP) [2], [9], [11] and the generalized memory polynomial (GMP) [2], [11]–[13], both of which can be interpreted to be special cases of the Volterra series [2], [14]–[16]. Such approaches allow for efficient direct and inverse modeling of nonlinear systems with memory, while also supporting straight-forward parameter estimation, through, e.g., linear least-squares (LS), as they are known to be linear-in-parameters models [11]. However, the processing complexity per linearized sample is also relatively high, particularly with GMP and other more complete Volterra series type of approaches, though also several works exist where complexity reduction is pursued [15], [17]–[21]. Specifically, the works in [18], [19], [22] present predistorter and PA modeling methods that build on spline-based basis functions – an approach that is technically considered also in this article, in the form of spline-interpolated lookup tables (LUTs).

In this paper, we develop and describe two new DPD solutions whose linearization capabilities are similar to those of the well-established polynomial-based solutions, while at the same time offering a substantially reduced DPD main path processing and parameter learning complexities. The development of such reduced-complexity DPD solutions is mainly motivated by the following four facts or tendencies. First, the channel bandwidths in NR are substantially larger than those in LTE-based systems. Specifically, up to 100 MHz and 400 MHz continuous channel bandwidths are already specified in NR Release-15 at frequency range 1 (FR-1; below 6 GHz bands) and FR-2 (24-40 GHz bands), respectively, [23], which imply increased DPD processing rates. Second, the actual unwanted emission requirements, particularly in the form of total radiated power (TRP) based adjacent channel leakage ratio (ACLR), are largely relaxed in NR FR-2 systems, being only in the order of 26-28 dB [23], increasing the feasibility of simplified DPD solutions. Third, the medium-range and local area base-stations (BSs) adopt substantially reduced transmit powers [23], compared to classical macro base-stations, hence the available power budget of the DPD solutions is also reduced. Finally, as observed recently in [5], even continuous learning may be needed at FR-2 and other mmWave active array systems, hence developing methods which reduce the parameter learning complexity becomes important.

The first new DPD method proposed in this paper, referred to as the spline-based Hammerstein (SPH) approach, builds on complex spline-interpolated LUT followed by a linear finite impulse response (FIR) filter. The interpolation allows to use a small amount of points in the LUT, while the linear filter facilitates basic memory modeling. Gradient-based

P. Pascual Campo, A. Brihuega, L. Anttila, M. Turunen, M. Allén, and M. Valkama are with the Department of Electrical Engineering, Tampere University, Tampere, Finland. e-mail: pablo.pascualcampo@tuni.fi

D. Korpi is with Nokia Bell Labs, Espoo, Finland.

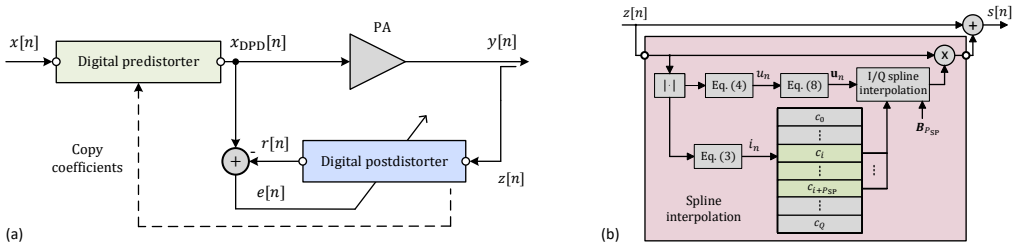


Fig. 1. Illustration of a) the considered DPD system building on indirect learning architecture (ILA), and b) the injection-based complex spline-interpolated LUT scheme utilized inside the proposed digital predistorter and digital postdistorter entities.

learning algorithms are also derived, to efficiently estimate the LUT control points as well as the linear filter parameters in a decoupled manner. The second proposed DPD method, referred to as the spline-based memory polynomial (SMP), consists of multiple parallel spline-interpolated LUTs and an input delay line such that more versatile memory modeling can be achieved when summing together the outputs of the parallel LUTs. Through spline interpolation, the size of all parallel LUTs can be kept small, while gradient-adaptive learning rule is again derived to estimate the control points of the involved parallel LUTs. For both proposed models – the SPH DPD and the SMP DPD – comprehensive computational complexity analyses are provided, while also comparing to ordinary gradient-adaptive canonical MP DPD system. Then, extensive RF measurement results are provided, covering several different FR-1 PA samples, channel bandwidth cases as well as base-station classes. Additionally, a state-of-the-art 28 GHz active antenna array, specifically Anokiwave AWMF-0129, is successfully linearized with 100 MHz and 200 MHz 5G NR channel bandwidths.

In general, it is noted that LUT-based PA linearization is, as such, a well-known approach, see, e.g., [8], [24]–[27] and the references therein. However, the PA memory aspects are not considered in [24], while fairly sizeable LUTs without interpolation are considered in [8], [26]. Additionally, a linearly-interpolated LUT-type implementation of a memory polynomial is described in [25] while the learning is based on classical LS model fitting. Furthermore, in [27], a DPD structure that includes two parallel Hammerstein systems compensating for the PA AM-AM and AM-PM responses, with Catmull-Rom spline interpolation, is presented. The model identification is based on a separable LS technique, specifically using a Levenberg-Marquardt algorithm to identify the DPD coefficients. The main path and training complexities are thus high when compared to the methods presented in this article. It is finally also noted that multi-dimensional LUT based solutions exist [28]–[30]. However, the LUT size in the nested LUT scheme in [28] grows exponentially with the memory depth, thus requiring unfeasible total LUT size when the linearized system exhibits substantial memory. The 2-dimensional LUT technique in [29] is, in turn, limited in its memory modeling capability, since it uses a weighted average of past amplitude samples to index the second LUT dimension. Finally, [30] combines an orthogonal matching pursuit algorithm to select the best LUTs in the forward path, and a partial least squares algorithm to estimate the DPD coefficients, which usually involves high complexity mainly due to the matrix inversion in the LS problem.

In the DPD system context of Fig. 1, the novelty and contributions of this article can be summarized as follows:

- New linear-in-parameters formulation for utilizing spline-interpolated I/Q LUTs in DPD systems, incorporating also the so-called injection-based DPD structure, is provided;
- New Hammerstein DPD solution utilizing the spline-interpolated I/Q LUT and decoupled gradient-based learning is proposed and derived;
- New memory polynomial DPD solution utilizing multiple parallel spline-interpolated I/Q LUTs and gradient-based learning is proposed and derived;
- Comprehensive computational complexity analysis of the methods is provided;
- Extensive performance-complexity assessments using versatile RF measurement examples at sub-6 GHz and 28 GHz bands are provided;

Compared to the existing literature, the new DPD formulation with spline-interpolated I/Q LUTs allows, in general, for (*i*) using any typical linear estimator (gradient or least-squares) to learn or update the LUT entries and (*ii*) reducing the main path processing complexity clearly when compared to ordinary canonical MP DPD. Specifically, the main path complexity and particularly the learning complexity are both reduced when compared to gradient-based canonical MP, owing to the use of the derived gradient-based learning in combination with the interpolated LUTs, since no basis function orthogonalization [7] nor self-orthogonalized learning procedure [25], [31] is needed with the proposed methods. Thus, by utilizing the proposed solutions, posing a remarkably low learning complexity, even continuous DPD adaptation/tracking is potentially viable.

The rest of the paper is organized as follows. Section II describes the I/Q spline interpolation scheme used throughout this paper, and presents the proposed SPH and SMP predistorter models. Section III derives and presents then the gradient-descent parameter learning algorithms for both DPD models. A complexity analysis and comparison of the proposed DPD solutions is provided in Section IV. Section V describes the RF measurement setups, and presents the corresponding measurement results and their analyses. Finally, conclusions are drawn in Section VI.

Throughout the rest of this article, matrices are denoted by

capital boldface letters, e.g., $\mathbf{A} \in \mathbb{C}^{M\times N}$, while vectors are denoted by lowercase boldface letters, e.g., $\mathbf{v} \in \mathbb{C}^{M\times 1} = [v_1\ v_2\ \cdots\ v_M]^T$. Ordinary transpose is represented as $(\cdot)^T$. Additionally, the complex conjugate, absolute value, floor, and ceil operators are represented as $(\cdot)^*$, $|\cdot|$, $\lfloor\cdot\rfloor$, and $\lceil\cdot\rceil$, respectively.

II. Proposed DPD Models

In this section, we introduce the proposed I/Q spline interpolation scheme, followed by the corresponding formulation of the SPH and SMP DPD models. For notational convenience, we formulate the mathematical presentation in the context of the indirect learning architecture (ILA) for postdistorter processing, with $z[n]$ and $r[n]$ denoting the postdistorter input and output, respectively. The input and output signals of the actual predistortion stage are $x[n]$ and $x_{\rm DPD}[n]$, as illustrated in Fig. 1.

A. Background and Basics

Building on piece-wise polynomials, spline based modeling and interpolation seeks to determine a smooth curve that approximates or conforms to a set of points, commonly known as control points [32]. Consequently, the input signal range is divided into several pieces, and the polynomials model the nonlinear system behavior in the corresponding regions under continuity and smoothness constraints. With this approach, simple low-order functions can be adopted, per region, in contrast to methods where a single high-order function or polynomial models the whole input range.

Traditionally, spline modeling has been applied to real-valued signals and systems [32]–[35]. However, in the context of radio communications, complex I/Q signals are utilized, and therefore the spline models need to be extended to the complex domain. Specifically, in this paper, we consider complex baseband models of RF nonlinearities, particularly those stemming from PA, for DPD purposes. To first shortly illustrate how splines can be applied to RF nonlinearity modelling at baseband, we start with the well-known memoryless polynomial, written for an arbitrary input signal $x_{\rm in}[n]$ as

$$x_{\rm out}[n] = \sum_{p=1,\ p\,\text{odd}}^{P} \alpha_p x_{\rm in}[n]|x_{\rm in}[n]|^{p-1}, \tag{1}$$

where $\alpha_p \in \mathbb{C}$ are the corresponding polynomial coefficients [9], [12]. Setting $\alpha_1 = 1$, without loss of generality, this can be re-written as

$$\begin{aligned} x_{\rm out}[n] &= x_{\rm in}[n](1 + \alpha_3|x_{\rm in}[n]|^2 + \cdots + \alpha_P|x_{\rm in}[n]|^{P-1}) \\ &= x_{\rm in}[n](1 + F(|x_{\rm in}[n]|)), \end{aligned} \tag{2}$$

where the function $F(\cdot) = F_{\rm I}(\cdot) + jF_{\rm Q}(\cdot)$ is a real-to-complex mapping. Thus, the baseband equivalent nonlinearity model consists of two real-valued functions $F_{\rm I}(\cdot)$ and $F_{\rm Q}(\cdot)$, both dependent only on the absolute value of the input signal.

B. Proposed I/Q Spline Interpolation Scheme

In general, the above model structure shown in (2) can be used for both PA direct modeling as well as PA inverse modeling, i.e., DPD. In the context of DPD, the nonlinear functions can be implemented efficiently with, for example, LUTs.

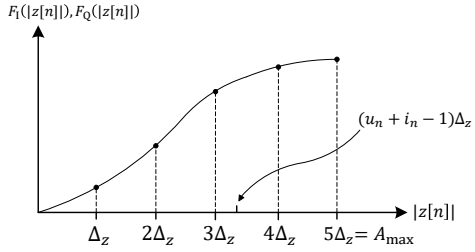


Fig. 2. Conceptual illustration of the nonlinear model regions of $F_{\rm I}(|z[n]|)$ and $F_{\rm Q}(|z[n]|)$ with respect the input magnitude $|z[n]|$ assuming $K = 5$ regions. Also an example of the input envelope value is shown, in this particular case within region $i_n = 4$, where $u_n \in [0, \Delta_z)$ denotes the normalized input envelope within a region.

To formulate a LUT-based DPD system in the linearization context of Fig. 1, we describe in this article spline-interpolated LUTs, i.e., small LUTs with spline interpolation to obtain the intermediate values. By adopting the notations in Fig. 1, such spline-based modeling of the nonlinear functions $F_{\rm I}(\cdot)$ and $F_{\rm Q}(\cdot)$ is illustrated at conceptual level in Fig. 2, where the input is a unipolar signal $|z[n]|$ with a maximum amplitude of $A_{\max}$. We adopt uniform equi-spaced splines with knot spacing (region width) of $\Delta_z > 0$, thus resulting in a total of $K = A_{\max}/\Delta_z$ regions. These regions are built, and accessed at time instant n, through the span index i_n and abscissa value u_n, defined similar to [36] as

$$i_n = \left\lfloor \frac{|z[n]|}{\Delta_z} \right\rfloor + 1, \tag{3}$$

$$u_n = \frac{|z[n]|}{\Delta_z} - (i_n - 1). \tag{4}$$

Here, i_n denotes the index of the selected region at time instant n, and u_n, $0 \le u_n < \Delta_z$, represents the normalized value of the corresponding input envelope within the current region i_n.

In general, adopting uniform splines allows the spline-interpolated output signal to take a very simple form, discussed also in [27] in the context of real-valued systems. The outputs of the I and Q splines can now be written as

$$F_{\rm I}(|z[n]|) = \mathbf{g}_n^T \mathbf{c}^{\rm re}, \tag{5}$$

$$F_{\rm Q}(|z[n]|) = \mathbf{g}_n^T \mathbf{c}^{\rm im}, \tag{6}$$

where $\mathbf{c}^{\rm re}$ and $\mathbf{c}^{\rm im}$ contain the Q control points of each spline. The vector $\mathbf{g}_n \in \mathbb{R}^{Q\times 1}$, in turn, is defined as

$$\mathbf{g}_n = \begin{bmatrix} 0 & \cdots & 0 & \mathbf{u}_n^T \mathbf{B}_{P_{\rm SP}} & 0 & \cdots & 0 \end{bmatrix}^T, \tag{7}$$

where

$$\mathbf{u}_n = \begin{bmatrix} u_n^{P_{\rm SP}} & u_n^{P_{\rm SP}-1} & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^{(P_{\rm SP}+1)\times 1}, \tag{8}$$

and $\mathbf{B}_{P_{\rm SP}} \in \mathbb{R}^{(P_{\rm SP}+1)\times(P_{\rm SP}+1)}$ is the *spline basis matrix* of order $P_{\rm SP}$. In (7), the term $\mathbf{u}_n^T \mathbf{B}_{P_{\rm SP}}$ of size $1 \times (P_{\rm SP}+1)$ is located such that the starting index is i_n. Thus, at a given time instant n, only the control points $c_{i_n}, c_{i_n+1}, \ldots, c_{i_n+P_{\rm SP}}$ contribute to the output. It is noted that for simplicity, we

assume in this work that the spline order P_{SP} does not depend on the region.

Using (5) and (6), while following the model structure in (2), the complex-valued output of the instantaneous nonlinear system, $s[n]$, can be constructed as

$$\begin{aligned} s[n] &= z[n] + z[n]\mathbf{g}_n^T(\mathbf{c}^{\mathrm{re}} + j\mathbf{c}^{\mathrm{im}}) \\ &= z[n] + z[n]\mathbf{g}_n^T\mathbf{c}, \end{aligned} \tag{9}$$

where $\mathbf{c} \in \mathbb{C}^{Q\times 1} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{Q-1} \end{bmatrix}^T$ is the overall complex-valued LUT containing the control points for the I and Q components. The interpolation scheme is further detailed in Fig. 1(b). We also note that the total number of control points with K regions and spline interpolation order P_{SP} is $Q = K + P_{\mathrm{SP}}$.

Importantly, the spline output $\mathbf{g}_n^T\mathbf{c}$ in (9) is defined as a deviation from unit gain. We refer to such structure as an injection-based scheme. Specifically, with this formulation, if $\mathbf{c}$ is initialized as an all-zero vector, the nonlinear system output will be the original input signal, i.e. $s[n] = z[n]$. By following this formulation, e.g., the gain ambiguities between the nonlinear spline and a cascaded FIR filter can be effectively removed – an issue that is relevant in the following Hammerstein DPD system – as the linear filter alone will handle the gain in the system. Additionally, the number of required bits in $\mathbf{c}$ in a fixed-point implementation is generally reduced, as this formulation reduces its dynamic range.

C. Spline-Interpolated Hammerstein DPD

This subsection introduces the proposed SPH scheme which builds on a Hammerstein structure where the involved nonlinearity is modelled with a complex spline-interpolated LUT. Following the proposed interpolation scheme presented above, in (9), we thus express the output of the instantaneous nonlinear block in the Hammerstein structure as

$$s_{\mathrm{SPH}}[n] = z_{\mathrm{SPH}}[n] + z_{\mathrm{SPH}}[n]\mathbf{g}_n^T\mathbf{c}. \tag{10}$$

It is noted that the term $\mathbf{g}_n$ depends on the B-spline *basis matrix* $\mathbf{B}_{P_{\mathrm{SP}}}$. This matrix can be precomputed for the given type of splines and polynomial order, and can be therefore considered as static. As a concrete example, in this article we focus on 3$^{\mathrm{rd}}$ order ($P_{\mathrm{SP}} = 3$, cubic interpolation) B-splines, although other spline orders are tested and demonstrated as well. In this case, the basis matrix can be expressed as [33]

$$\mathbf{B}_3 = \frac{1}{6}\begin{bmatrix} \frac{-1}{\Delta_z^3} & \frac{3}{\Delta_z^3} & \frac{-3}{\Delta_z^3} & \frac{1}{\Delta_z^3} \\ \frac{3}{\Delta_z^2} & \frac{-6}{\Delta_z^2} & \frac{3}{\Delta_z^2} & 0 \\ \frac{-3}{\Delta_z} & 0 & \frac{3}{\Delta_z} & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}. \tag{11}$$

Next, after having derived the expression for the memoryless nonlinear signal model, the memory effects are incorporated through the FIR filter stage that is common to all regions. Hence, the overall output signal $r_{\mathrm{SPH}}[n]$ can be directly expressed as

$$r_{\mathrm{SPH}}[n] = \mathbf{h}^T\mathbf{s}_n, \tag{12}$$

where $\mathbf{h} \in \mathbb{C}^{M_{\mathrm{SPH}}\times 1} = \begin{bmatrix} h_0 & h_1 & \cdots & h_{M_{\mathrm{SPH}}-1} \end{bmatrix}^T$ contains the filter coefficients, with M_{SPH} denoting the number of taps in the model, while $\mathbf{s}_n \in \mathbb{C}^{M_{\mathrm{SPH}}\times 1} = \begin{bmatrix} s_{\mathrm{SPH}}[n] & s_{\mathrm{SPH}}[n-1] & \cdots & s_{\mathrm{SPH}}[n-M_{\mathrm{SPH}}+1] \end{bmatrix}^T$. The overall processing structure is illustrated in Fig. 3(a).

D. Spline-Interpolated MP LUT DPD

This subsection formulates the proposed SMP DPD model. Inspired by [8], a memory polynomial type parallel branched structure is adopted to model the memory effects, while the actual parallel nonlinearities are each implemented through the complex spline-interpolated LUTs presented above.

Following this approach, the proposed SMP processing can thus be expressed as

$$r_{\mathrm{SMP}}[n] = z_{\mathrm{SMP}}[n] + \sum_{m=0}^{M_{\mathrm{SMP}}-1} z_{\mathrm{SMP}}[n-m]\mathbf{g}_{n-m}^T\mathbf{q}_m, \tag{13}$$

where M_{SMP} denotes the considered memory order while $\mathbf{q}_m$, $m = 0, 1, \ldots, M_{\mathrm{SMP}} - 1$, are the M_{SMP} LUTs of the model, containing the control points for the spline interpolation in each parallel branch. The proposed SMP processing structure, adopting also the injection principle but in generalized form, is illustrated in Fig. 3(b).

Generally, in terms of the modeling capabilities, the SMP is a richer model compared to SPH, while it also naturally entails higher complexity. These models will be assessed and compared to classical DPD solutions in terms of complexity and performance in Sections IV and V.

III. Parameter Learning Rules

In this section, we derive efficient gradient-descent type learning rules for both proposed DPD approaches, to adaptively estimate and track the unknown parameters in each of the models. Notation-wise, to allow for sample-adaptive estimation, we denote the vectors to be estimated with a subindex n, i.e., $\mathbf{c}_n$ and $\mathbf{h}_n$ for SPH and $\mathbf{q}_{m,n}$, $m = 0, 1, \ldots, M_{\mathrm{SMP}}-1$, for SMP, to indicate their time-dependence.

A. SPH Learning Rules

To calculate the learning rule in the SPH case, the instantaneous error signal between $x_{\mathrm{DPD}}[n]$ and $r_{\mathrm{SPH}}[n]$, in the context of the considered ILA-type architecture is first extracted as

$$e_{\mathrm{SPH}}[n] = x_{\mathrm{DPD}}[n] - r_{\mathrm{SPH}}[n] = x_{\mathrm{DPD}}[n] - \mathbf{h}_n^T\mathbf{s}_n. \tag{14}$$

Then, to facilitate the gradient-descent learning [37], the cost function is defined as the instantaneous squared error, expressed as

$$J(\mathbf{h}_n, \mathbf{c}_n) = |e_{\mathrm{SPH}}[n]|^2. \tag{15}$$

The corresponding iterative learning rules are then obtained through the partial derivatives of $J(\mathbf{h}_n, \mathbf{c}_n)$ with respect to both parameter vectors to adapt, expressed formally as

$$\mathbf{h}_{n+1} = \mathbf{h}_n - \mu_{\mathrm{h}}[n]\nabla_{\mathbf{h}_n}J(\mathbf{h}_n, \mathbf{c}_n), \tag{16}$$

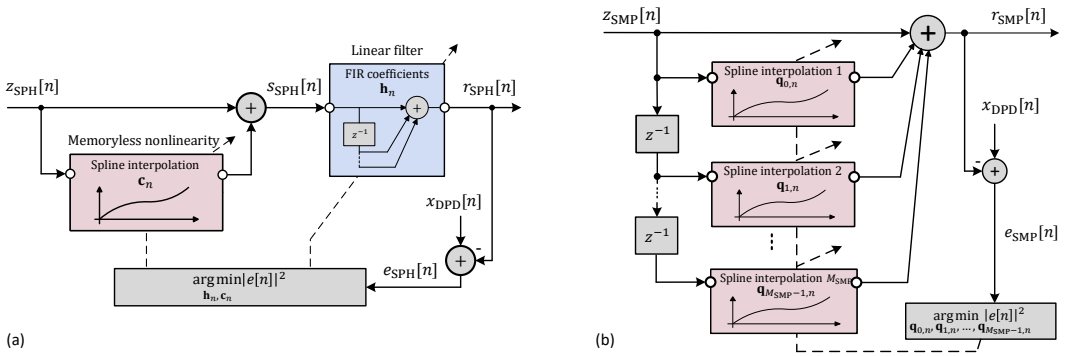


Fig. 3. Illustration of the postdistorter structures of a) the proposed SPH DPD system, and b) the proposed SMP DPD model. The *Spline interpolation* block(s) in both models comprises the scheme shown in Fig. 1(b).

$$\mathbf{c}_{n+1} = \mathbf{c}_n - \mu_{\mathrm{c}}[n]\nabla_{\mathbf{c}_n} J(\mathbf{h}_n, \mathbf{c}_n), \tag{17}$$

where $\nabla_{\mathbf{x}}$ refers to the complex gradient operator [37], [38] of a real-valued function against complex-valued parameter vector $\mathbf{x}$. Additionally, $\mu_{\mathrm{h}}[n]$ and $\mu_{\mathrm{c}}[n]$ are the learning rates for $\mathbf{h}_n$ and $\mathbf{c}_n$, respectively, at time instant n. After relatively straight-forward derivations, the resulting concrete learning rules read

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \mu_{\mathrm{h}}[n] e_{\mathrm{SPH}}[n] \mathbf{s}_n^*, \tag{18}$$

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu_{\mathrm{c}}[n] e_{\mathrm{SPH}}[n] \mathbf{\Sigma}_n^T \mathbf{Z}_n^* \mathbf{h}_n^*, \tag{19}$$

where the diagonal matrix $\mathbf{Z}_n \in \mathbb{C}^{M_{\mathrm{SPH}} \times M_{\mathrm{SPH}}} = \mathrm{diag}\left\{z_{\mathrm{SPH}}[n],\ z_{\mathrm{SPH}}[n-1],\ \cdots,\ z_{\mathrm{SPH}}[n-M_{\mathrm{SPH}}+1]\right\}$, and $\mathbf{\Sigma}_n$ contains M_{SPH} previous instances of $\mathbf{g}_n$, defined as $\mathbf{\Sigma}_n \in \mathbb{R}^{M_{\mathrm{SPH}} \times Q} = \begin{bmatrix}\mathbf{g}_n & \mathbf{g}_{n-1} & \cdots & \mathbf{g}_{n-M_{\mathrm{SPH}}+1}\end{bmatrix}^T$. These learning rules in (18) and (19) are executed in parallel such that both parameter vectors are updated simultaneously. For readers' convenience, an example illustration of the structure of the matrix $\mathbf{\Sigma}_n$ is given in (20), for $M_{\mathrm{SPH}} = 4$, $Q = 9$, and $P_{\mathrm{SP}} = 3$, assuming representative example values of the index variable i_n.

$$\mathbf{\Sigma}_n = \begin{bmatrix} 0 & 0 & \underbrace{[* & * & * & *]}_{\mathbf{u}_n^T\mathbf{B}_{P_{\mathrm{SP}}}} & 0 & 0 & 0 & \\ \underbrace{[* & * & * & *]}_{\mathbf{u}_{n-1}^T\mathbf{B}_{P_{\mathrm{SP}}}} & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & \underbrace{[* & * & * & *]}_{\mathbf{u}_{n-2}^T\mathbf{B}_{P_{\mathrm{SP}}}} & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \underbrace{[* & * & * & *]}_{\mathbf{u}_{n-3}^T\mathbf{B}_{P_{\mathrm{SP}}}} & \end{bmatrix} \begin{matrix} i_n = 3 \\ i_{n-1} = 1 \\ i_{n-2} = 4 \\ i_{n-3} = 6 \end{matrix} \tag{20}$$

Note that the term $\mathbf{u}_n^T\mathbf{B}_{P_{\mathrm{SP}}}$ is located in $\mathbf{\Sigma}_n$ at each iteration n according to the span index i_n, as shown in (7). It is noted that the derived learning rules in (18) and (19) are novel, as the overall Hammerstein system is known to be not linear in its parameters.

B. SMP Learning Rules

We next derive gradient-based iterative learning rules for the SMP model. Different to SPH case, the SMP model does not contain cascaded filters while the learning entity considers the M_{SMP} parallel spline-interpolated LUTs, specifically their control points $\mathbf{q}_{m,n}$, $m = 0, 1, \ldots, M_{\mathrm{SMP}} - 1$.

Following now a similar approach as earlier, the instantaneous error signal is first defined in the context of ILA-based learning as

$$\begin{aligned} e_{\mathrm{SMP}}[n] &= x_{\mathrm{DPD}}[n] - r_{\mathrm{SMP}}[n] \\ &= x_{\mathrm{DPD}}[n] - z_{\mathrm{SMP}}[n] - \sum_{m=0}^{M_{\mathrm{SMP}}-1} z_{\mathrm{SMP}}[n-m]\mathbf{g}_{n-m}^T\mathbf{q}_{m,n}. \end{aligned} \tag{21}$$

For the gradient-descent learning, the cost function is defined as a function of the instantaneous error signal as

$$J(\mathbf{q}_{0,n}, \mathbf{q}_{1,n}, \cdots, \mathbf{q}_{M_{\mathrm{SMP}}-1,n}) = |e_{\mathrm{SMP}}[n]|^2. \tag{22}$$

Then, by adopting again the complex gradient operator [38], the learning rule for the mth LUT can be written as

$$\mathbf{q}_{m,n+1} = \mathbf{q}_{m,n} - \mu_{\mathrm{q}_m} \nabla_{\mathbf{q}_{m,n}} J(\mathbf{q}_{0,n}, \mathbf{q}_{1,n}, \cdots, \mathbf{q}_{M_{\mathrm{SMP}}-1,n}), \tag{23}$$

while by following the complex differentiation steps, the final learning rule for the mth LUT reads

$$\mathbf{q}_{m,n+1} = \mathbf{q}_{m,n} + \mu_{\mathrm{q}_m}[n]\ e_{\mathrm{SMP}}[n]\ z_{\mathrm{SMP}}^*[n-m]\ \mathbf{g}_{n-m}. \tag{24}$$

These learning rules are adopted for all the involved M_{SMP} LUTs in parallel.

IV. Computational Complexity Analysis

In this section, a computational complexity analysis and comparison between the proposed SPH, SMP and a widely-applied canonical MP DPD with self-orthogonalizing least-mean square (LMS) [37] parameter adaptation is presented. LMS type adaptation is deliberately assumed also for MP DPD, for the fairness of the comparison. The complexity analysis is carried out in terms of real multiplications per

linearized data sample, as multiplications are commonly more resource-intensive operations than additions in digital signal processing (DSP) implementations [11].

The quantitative complexity assessment of the proposed gradient-adaptive SPH DPD and SMP DPD follows the exact processing steps described in Sections II and III. It is noted that the complexity expressions reported below basically represent an upper bound for the required arithmetical operations, as in real implementations some elementary or trivial operations such as multiplying by any integer power of 2 or $1/2$ does not really reflect any actual complexity, while are included as normal operations in the expressions for simplicity. Additionally, it is noted that the modulus operator, needed in (3) and (4), is assumed to be calculated with the alpha max beta min algorithm [39]. Finally, in the complexity analysis, we consider uniform splines with $\Delta_z = \Delta_x = 1$.

A. Complexity of Proposed SPH Method

With reference to Fig. 3(a) and the underlying processing elements, the generic complexity expressions can be stated in a straight-forward manner as follows:

- *DPD main path*, starting with the input signal $x[n]$. The complexity of the predistorter intermediate signal, $s_{\mathrm{DPD}}[n]$, includes the processing in (3), (4), and (10) but with $x[n]$ as the input. These together with the FIR filtering in (12) to calculate $x_{\mathrm{DPD}}[n]$ yield the following complexity expressions
  1) $s_{\mathrm{DPD}}[n] \rightarrow P_{\mathrm{SP}}^2 + 4P_{\mathrm{SP}} + 10$.
  2) $x_{\mathrm{DPD}}[n] \rightarrow 4M_{\mathrm{SPH}}$.
- *DPD learning*, for observed signal $z_{\mathrm{SPH}}[n]$. The generation of the error signal $e_{\mathrm{SPH}}[n]$ contains the same multiplication operations as in $s_{\mathrm{DPD}}[n]$ and $x_{\mathrm{DPD}}[n]$, due to the ILA architecture. The complexity of updating $\mathbf{h}_n$ and $\mathbf{c}_n$ corresponds to calculating (18) and (19), respectively. Overall, we thus get
  1) $e_{\mathrm{SPH}}[n] \rightarrow P_{\mathrm{SP}}^2 + 4P_{\mathrm{SP}} + 4M_{\mathrm{SPH}} + 10$.
  2) $\mathbf{h}_{n+1} \rightarrow 4M_{\mathrm{SPH}} + 2$.
  3) $\mathbf{c}_{n+1} \rightarrow 2P_{\mathrm{SP}}M_{\mathrm{SPH}} + 4P_{\mathrm{SP}} + 6M_{\mathrm{SPH}} + 6$.

Interestingly, it is noted that the amount of multiplications in the DPD main path does not depend on the chosen number of control points Q, or equivalently the number of regions, as the spline-interpolation algorithm basically utilizes $P_{\mathrm{SP}}+1$ control points for any given region.

B. Complexity of Proposed SMP Method

With the SMP approach, as shown in (13) for post-distortion, there is no separate linear filtering stage but the overall DPD output is composed as a sum of M_{SMP} parallel spline-interpolated LUTs with input samples $x[n-m]$, $m = 0, 1, \ldots, M_{\mathrm{SMP}}-1$. Therefore, with reference to Fig. 3(b) and the underlying processing ingredients described in Sections II and III, the main path and parameter learning complexities can be stated as follows:

- *DPD main path*, starting with the input signal $x[n]$. The complexity involves calculating $x_{\mathrm{DPD}}[n]$, as in (13), with $x[n]$ as the input. By taking into account that at time instant n, only $\mathbf{g}_n$ needs to be calculated while $\mathbf{g}_{n-1}, \ldots, \mathbf{g}_{n-M_{\mathrm{SMP}}+1}$ are available from previous sample instant, we obtain the following overall complexity expression
  1) $x_{\mathrm{DPD}}[n] \rightarrow P_{\mathrm{SP}}^2 + 3P_{\mathrm{SP}} + 2P_{\mathrm{SP}}M_{\mathrm{SMP}} + 6M_{\mathrm{SMP}} + 4$.
- *DPD learning path*, for observed signal $z_{\mathrm{SMP}}[n]$. Due to the ILA architecture, the involved complexity of calculating the error signal $e_{\mathrm{SMP}}[n]$ is, arithmetically, the same as calculating $x_{\mathrm{DPD}}[n]$. Additionally, the complexity of updating one of the LUTs or spline control point vectors, $\mathbf{q}_{m,n}$, corresponds to calculating (24). Thus, we get
  1) $e_{\mathrm{SMP}}[n] \rightarrow P_{\mathrm{SP}}^2 + 3P_{\mathrm{SP}} + 2P_{\mathrm{SP}}M_{\mathrm{SMP}} + 6M_{\mathrm{SMP}} + 4$.
  2) $\mathbf{q}_{m,n+1} \rightarrow 2P_{\mathrm{SP}} + 8$.

C. Complexity of Reference MP DPD

When considering the LMS-adaptive MP DPD with monomial basis functions (BFs), in the context of ILA architecture in Fig. 1(a), we first write the postdistorter output sample as

$$r_{\mathrm{MP}}[n] = \mathbf{w}_n^T \mathbf{l}_n, \tag{25}$$

where $\mathbf{w}_n \in \mathbb{C}^{m\times 1}$ is the MP DPD coefficient vector, with $m = \lceil \frac{P_{\mathrm{MP}}}{2} \rceil M_{\mathrm{MP}}$ denoting the number of coefficients, while P_{MP} and M_{MP} are the assumed polynomial order and memory length (per nonlinearity order), respectively. Additionally, the vector of the basis function samples $\mathbf{l}_n$ used to calculate the current output is as defined in (26), next page, where $z_{\mathrm{MP}}[n]$ denotes the observed feedback signal at postdistorter input.

Once $r_{\mathrm{MP}}[n]$ is calculated, the error signal can be directly obtained as

$$e_{\mathrm{MP}}[n] = x_{\mathrm{DPD}}[n] - r_{\mathrm{MP}}[n] = x_{\mathrm{DPD}}[n] - \mathbf{w}_n^T \mathbf{l}_n, \tag{27}$$

and the coefficient update can be written as

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_{\mathrm{w}}[n] e_{\mathrm{MP}}[n] \mathbf{R}^{-1} \mathbf{l}_n^*, \tag{28}$$

where $\mu_{\mathrm{w}}[n]$ is the learning rate, and $\mathbf{R}^{-1}$ is the inverse of the autocorrelation matrix of the PA output basis function samples [37]. We assume that a block of N_{B} samples is used to calculate the sample estimate of $\mathbf{R}$, and include below the corresponding complexity for completeness of the study. Importantly, it is also noted that the self-orthogonalizing type transformation $\mathbf{R}^{-1}$ in (28) is an important ingredient for stable operation, as the MP basis functions in (26) are known to be largely correlated [40]. Alternatively, orthogonal polynomial type set of basis functions could be used [40], [41], though with increased main path complexity. The SPH and SMP DPD related learning rules in (18)-(19) and (24), on the other hand, do not suffer from such correlation challenge, and are shown in Section V to provide reliable linearization without any additional (self-)orthogonalization. This is one clear benefit, complexity-wise, compared to the existing gradient-adaptive reference DPD solutions.

Building on above, the self-orthogonalizing LMS-adaptive MP DPD complexity can be detailed as follows

$$\mathbf{l}_n = \Big[z_{\mathrm{MP}}[n] \quad z_{\mathrm{MP}}[n]\,|z_{\mathrm{MP}}[n]|^2 \;\cdots\; z_{\mathrm{MP}}[n]\,|z_{\mathrm{MP}}[n]|^{P_{\mathrm{MP}}-1} \quad z_{\mathrm{MP}}[n-1] \quad z_{\mathrm{MP}}[n-1]\,|z_{\mathrm{MP}}[n-1]|^2 \;\cdots\; z_{\mathrm{MP}}[n-1]\,|z_{\mathrm{MP}}[n-1]|^{P_{\mathrm{MP}}-1}$$
$$z_{\mathrm{MP}}[n-M_{\mathrm{MP}}+1] \quad z_{\mathrm{MP}}[n-M_{\mathrm{MP}}+1]\,|z_{\mathrm{MP}}[n-M_{\mathrm{MP}}+1]|^2 \;\cdots\; z_{\mathrm{MP}}[n-M_{\mathrm{MP}}+1]\,|z_{\mathrm{MP}}[n-M_{\mathrm{MP}}+1]|^{P_{\mathrm{MP}}-1} \Big]^T. \tag{26}$$

TABLE I
COMPLEXITY EXPRESSIONS IN TERMS OF REAL MULTIPLICATIONS PER SAMPLE FOR THE PROPOSED SPH, THE PROPOSED SMP AND THE REFERENCE CANONICAL MP METHODS, COVERING BOTH THE DPD MAIN PATH PROCESSING AND THE DPD PARAMETER LEARNING, WITH $m = \lceil \frac{P_{\mathrm{MP}}}{2} \rceil M_{\mathrm{MP}}$

Operation		**SPH model**	**SMP model**	**MP model**
Predistortion	Nonlinearity	$P_{\mathrm{SP}}^2 + 4P_{\mathrm{SP}} + 10$	$P_{\mathrm{SP}}^2 + 3P_{\mathrm{SP}} + 2P_{\mathrm{SP}}M_{\mathrm{SMP}} + 6M_{\mathrm{SMP}} + 4$	$3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil - 2$
	Filtering	$4M_{\mathrm{SPH}}$	0	$4m$
	Total	$P_{\mathrm{SP}}^2 + 4P_{\mathrm{SP}} + 4M_{\mathrm{SPH}} + 10$	$P_{\mathrm{SP}}^2 + 3P_{\mathrm{SP}} + 2P_{\mathrm{SP}}M_{\mathrm{SMP}} + 6M_{\mathrm{SMP}} + 4$	$3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil + 4m - 2$
Learning	Error signal	$P_{\mathrm{SP}}^2 + 4P_{\mathrm{SP}} + 4M_{\mathrm{SPH}} + 10$	$P_{\mathrm{SP}}^2 + 3P_{\mathrm{SP}} + 2P_{\mathrm{SP}}M_{\mathrm{SMP}} + 6M_{\mathrm{SMP}} + 4$	$3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil + 4m - 2$
	Update	$2P_{\mathrm{SP}}M_{\mathrm{SPH}} + 4P_{\mathrm{SP}} + 10M_{\mathrm{SPH}} + 8$	$M_{\mathrm{SMP}}(2P_{\mathrm{SP}} + 8)$	$4m^2 + 4m + 2$
	Total	$P_{\mathrm{SP}}(P_{\mathrm{SP}} + 2M_{\mathrm{SPH}} + 8) + 14M_{\mathrm{SPH}} + 18$	$P_{\mathrm{SP}}^2 + 3P_{\mathrm{SP}} + 4P_{\mathrm{SP}}M_{\mathrm{SMP}} + 14M_{\mathrm{SMP}} + 4$	$3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil + 4m^2 + 8m$

- DPD main path, starting with the input signal $x[n]$, in terms of real multiplications per linearized sample:
  1) MP BF samples $\rightarrow 3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil - 2$.
  2) $x_{\mathrm{DPD}}[n] \rightarrow 4m$.
- DPD training, for observed signal $z_{\mathrm{MP}}[n]$:
  1) MP BF samples $\rightarrow 3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil - 2$.
  2) $\mathbf{R}^{-1} \rightarrow m^3$.
  3) $r_{\mathrm{MP}}[n] \rightarrow 4m$.
  4) $e_{\mathrm{MP}}[n] \rightarrow 3\lceil \frac{P_{\mathrm{MP}}}{2} \rceil + 4m - 2$.
  5) $\mathbf{w}_{n+1} \rightarrow 4m^2 + 4m + 2$.

TABLE II
NUMERICAL COMPLEXITY VALUES, IN TERMS OF REAL MULTIPLICATIONS PER SAMPLE, FOR $P_{\mathrm{SP}} = 3$, $M_{\mathrm{SPH}} = M_{\mathrm{SMP}} = 4$, $Q_{\mathrm{SPH}} = Q_{\mathrm{SMP}} = 7$, $P_{\mathrm{MP}} = 11$, AND $M_{\mathrm{MP}} = 4$.

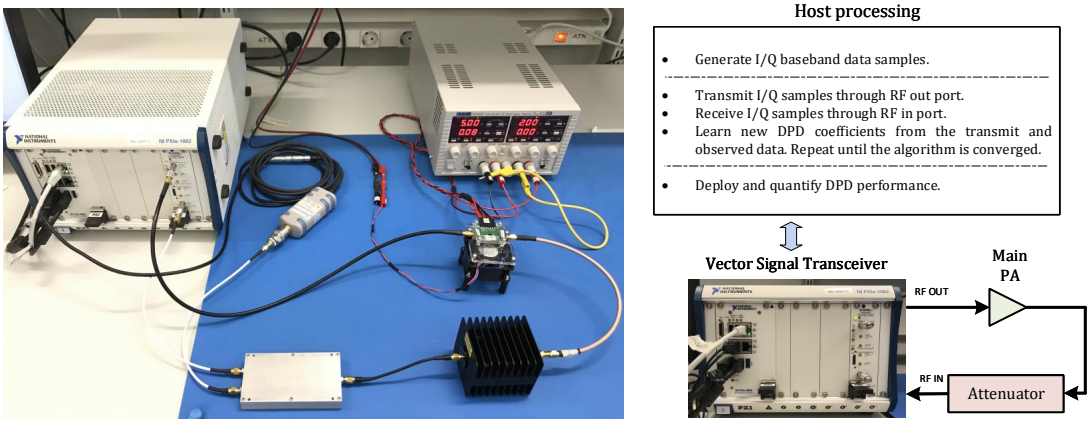
	SPH model	**SMP model**	**MP model**
No. of coefficients	14	31	24
Nonlinearity	24	63	16
Filtering	16	0	96
Total main path	40	63	112
Relative complexity wrt. MP	35.7%	56.3%	100%
Error signal	40	63	112
Coeff. update	84	56	2402
Total learning	124	119	2514
Relative complexity wrt. MP	4.9%	4.7%	100%

D. Summary and Comparison

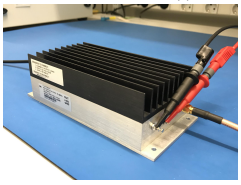
Table I collects and summarizes the deduced expressions for the numbers of real multiplications per sample needed for the fundamental main path processing and parameter learning stages in the proposed SPH, SMP and the reference MP DPD methods. In this table, when it comes to MP DPD, we have excluded the complexity related to the calculation of the elements of $\mathbf{R}$ and its inverse, as those are something that can be considered carried out offline, or within the very first phases of the overall learning procedure.

Next, to obtain concrete numerical complexity numbers and to carry out a comparison, we study an example case where the SPH and SMP DPD spline polynomial order is $P_{\mathrm{SP}} = 3$. Additionally, the number of control points per LUT is chosen to be $Q = 7$ for both SPH and SMP models, and the considered memory length is $M_{\mathrm{SPH}} = M_{\mathrm{SMP}} = 4$. These constitute a total number of 14 free parameters to be estimated in the SPH model and 31 free parameters in the SMP case. Then, the MP DPD polynomial order is chosen as $P_{\mathrm{MP}} = 11$, and the considered memory length per filter is $M_{\mathrm{MP}} = 4$. This configuration leads to 24 free parameters in the MP DPD. Similar type parametrizations are used also in the actual DPD measurements and experiments, in Section V.

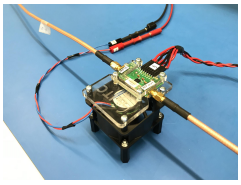
The resulting exact numerical processing complexities, expressed in terms of real multiplications per linearized sample, are presented in Table II. In these numerical values, when it comes to the SPH and SMP DPD, we have excluded the trivial operations, i.e., multiplications by zeros, ones and integer powers of two or half, stemming from the structure of $\mathbf{B}_3$ in (11). Overall, the results in Table II demonstrate the large complexity reduction provided by the proposed spline-based DPD approaches, with respect to the reference MP model, the relative complexities being 35.7% (SPH) and 56.3% (SMP) in the DPD main path to predistort the input signal. Furthermore, the required parameter learning complexity is also very remarkably reduced, with the relative complexity compared to the MP model being only approximately 5% in both SPH and SMP cases. This indicating that solutions like these might already facilitate even continuous DPD learning or parameter tracking in selected applications. Such continuous DPD learning can be important, e.g., in mmWave active array linearization where the effective nonlinear characteristics are known to be beam-dependent due to the load modulation phenomenon [5]. Additionally, owing to the largely reduced learning complexity, the feasibility of implementing both the



(a) RF measurement setup at FR-1.



(b) General purpose wideband PA.



(c) Skyworks NR Band 78 PA.

Fig. 4. Overall RF measurement setup at FR-1 and the sub-6 GHz PA modules used in the Experiments 1-2.

DPD parameter learning as well as the main path processing in the same chip increases.

V. Experimental Results

In order to evaluate and validate the proposed DPD concepts, three separate linearization experiments are carried out. The first two measurement scenarios are conducted with FR-1 (sub-6 GHz) PAs, including a general purpose wideband PA and a 5G NR Band 78 small-cell BS PA. The third experiment is then related to FR-2 and over-the-air (OTA) measurements where a state-of-the-art 28 GHz active antenna array with 64 integrated PAs and antenna units is linearized. For complexity assessment, we use the derived results in Table I, while again exclude the trivial operations, i.e., multiplications by zeros, ones and integer powers of two or half, stemming from the structure of the B-spline basis matrix $\mathbf{B}_{P_{\mathrm{SP}}}$. Additionally, we also provide the corresponding amounts of floating point operations (FLOPs) per sample. One complex multiplication is assumed to cost 6 FLOPs, while one complex-real multiplication and one complex sum both cost 2 FLOPs [42].

A. FR-1 Measurement Environment and Figures of Merit

The FR-1 measurement setup utilized for the first two experiments is illustrated in Fig. 4(a), and consists of a National Instruments PXIe-5840 vector signal transceiver (VST), facilitating arbitrary waveform generation and analysis between 0–6 GHz with instantaneous bandwidth of 1 GHz. This instrument is used as both the transmitter and the observation receiver, and includes also an additional host-processing based computing environment where all the digital waveform and DPD processing can be executed. In a typical conducted measurement, the baseband complex I/Q waveform is generated by MATLAB in the VST host environment, and fed to the device under test (DUT) through the VST transmit chain. The DUT output is then observed via the VST receiver, through an external attenuator. All DPD parameter learning and actual DPD main math processing stages are executed in the host environment. Finally, the actual DPD performance measurements are carried out where different random modulating data is used, compared to the learning phase.

As the DPD system figures of merit, we adopt the well-established error vector magnitude (EVM) and ACLR metrics, as defined for 5G NR in [23]. The EVM focuses on the passband transmit signal quality, and is defined as

$$\mathrm{EVM}\ (\%) = \sqrt{\frac{P_{\mathrm{error,\ eq.}}}{P_{\mathrm{ref.}}}} \times 100, \tag{29}$$

where $P_{\mathrm{error,\ eq.}}$ denotes the power of the error signal calculated between the ideal subcarrier symbols and the corresponding observed subcarrier samples at the PA output after zero forcing equalization removing the effects of the possible linear distortion [23]. Furthermore, $P_{\mathrm{ref.}}$ denotes the corresponding power of the ideal (reference) symbols. The ACLR is defined as the ratio of the transmitted power within the desired channel ($P_{\mathrm{desired\ ch}}$) and that in the left or right adjacent channel

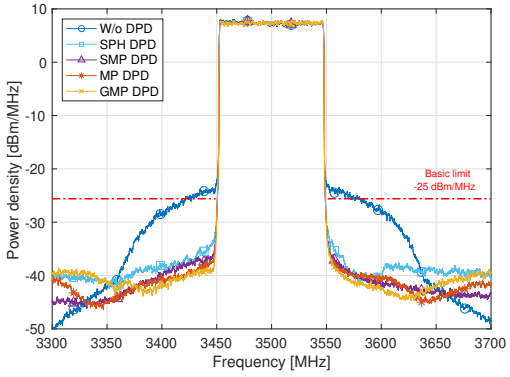


Fig. 5. Example illustration of linearization results in Experiment 1 (general purpose PA measured at 3.5 GHz), with 100 MHz channel bandwidth and PA output power of +27 dBm, while adopting $P_{\rm SP} = 3$, $Q_{\rm SPH} = Q_{\rm SMP} = 7$, $M_{\rm SPH} = 3$, $M_{\rm SMP} = 4$, $P_{\rm MP} = P_{\rm GMP} = 11$, $M_{\rm MP} = M_{\rm GMP} = 4$, and $G = 9$.

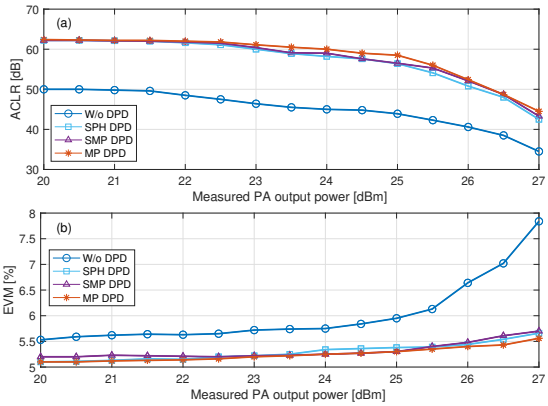


Fig. 6. Measured ACLR and EVM performance in Experiment 1 as a function of the PA output power, while adopting $P_{\rm SP} = 3$, $Q_{\rm SPH} = Q_{\rm SMP} = 7$, $M_{\rm SPH} = 3$, $M_{\rm SMP} = 4$, $P_{\rm MP} = 11$, and $M_{\rm MP} = 4$.

($P_{\text{adj. ch.}}$), expressed as

$$\text{ACLR (dB)} = 10\log_{10}\frac{P_{\text{desired ch.}}}{P_{\text{adj. ch.}}}, \tag{30}$$

measuring thus the out-of-band performance. While ACLR is, by definition, a relative measure, an explicit out-of-band spectral density limit, in terms of dBm/MHz measured with a sliding 1 MHz window in the adjacent channel region, is also defined for certain base-station types [23], referred to as the absolute basic limit in 3GPP terminology. Thus, the PA output spectral density in dBm/MHz is also quantified in the measurements, particularly in the context of local area and medium-range BS PAs [23]. Specifically, we quantify the maximum out-of-band spectral density, denoted as maximum dBm/MHz in the following experiments, which will allow for further comparison between the deployed DPD models.

All the forth-coming experiments utilize 5G NR Release-15 standard compliant OFDM downlink waveform and channel bandwidths [23], while the adopted carrier frequencies in each experiment are selected according to the available 5G NR bands and the available PA samples. In all experiments, the initial PAPR of the digital waveform is 9.5 dB, when measured at the 0.01% point of the instantaneous PAPR complementary cumulative distribution function (CCDF), and is then limited to 7 dB through well-known iterative clipping and filtering based processing, while also additional time-domain windowing is applied to suppress the inherent OFDM signal sidelobes. These impose an EVM floor of approximately 4% to the transmit signal. More specific waveform parameters such as the subcarrier spacing (SCS) and the occupied physical resource block (PRB) count are stated along the experiments.

Finally, in the upcoming FR-1 experiments, we also utilize LMS-adaptive GMP [12] as an additional reference DPD method for completeness. For simplicity, we assume the same amount of leading and lagging envelope delays [12], denoted here by G.

B. Experiment 1: General Purpose PA

The first experiment focuses on a general purpose wideband PA (Mini-Circuits ZHL-4240), illustrated in Fig. 4(b), as the actual amplification stage. The amplifier has a gain of 41 dB, and a 1-dB compression point of +31 dBm, being basically applicable in small-cell and medium-range base-stations. The transmit signal is a 5G NR downlink OFDM waveform, with 30 kHz subcarrier spacing and 264 active PRBs [23], yielding a passband width of 95.04 MHz. The RF center frequency is 3.5 GHz and the assumed channel bandwidth is 100 MHz. The I/Q samples are transmitted through the VST RF output port directly to the PA, facilitating a maximum output power of +27 dBm. The proposed and the reference DPD schemes are then adopted, and the performance quantification measurements are carried out. In all results, five ILA learning iterations are adopted while the signal length within each ILA iteration is 100,000 samples. In this experiment, the VST observation receiver runs at 491.52 MHz ($4\times$ oversampling).

Fig. 5 shows a snap-shot linearization example, at PA output power of +27 dBm, when $P_{\rm SP} = 3$ is chosen as the spline order in both the SPH and SMP models, while the number of control points is fixed to $Q_{\rm SPH} = Q_{\rm SMP} = 7$ and the memory filter orders are $M_{\rm SPH} = 3$ and $M_{\rm SMP} = 4$. Additionally, an LMS-based MP DPD ($P_{\rm MP} = 11$, $M_{\rm MP} = 4$) and an LMS-based GMP DPD [12] ($P_{\rm GMP} = 11$, $M_{\rm GMP} = 4$, $G = 9$) are also adopted and presented for reference. We can observe that the performances of the proposed SPH and SMP DPDs are very close to each other, and to those of the MP DPD and GMP DPD, despite the substantially reduced complexity. The figure also illustrates that all DPD methods basically satisfy the absolute basic limit requirement of -25 dBm/MHz, which is less severe than the classical 45 dB ACLR limit, and applies in medium-range BS cases with TX powers of higher than +24 dBm up to +38 dBm [23].

Fig. 6 then presents the behavior of the measured EVM and ACLR performance metrics, as functions of the PA output power, following the same DPD parameterization. Again, we can observe that the proposed SPH, SMP, and the MP DPD

TABLE III
SUMMARY OF DPD MAIN PATH PROCESSING COMPLEXITY AND LINEARIZATION PERFORMANCE IN EXPERIMENT 1, PA OUTPUT POWER IS +27 dBm.

	Running complexity								Model performance	
	P	**M**	**G**	**Q**	$\Delta_{z,x}$	**# of coefficients**	**FLOPs/sample**	**Mul./sample**	**EVM (%)**	**Max. dBm/MHz**
No DPD	-	-	-	-	-	-	-	-	7.82	-23.80
SPH DPD	2	3	-	7	1	12	55	28	5.61	-32.30
	3	3	-	7	1	13	69	36	5.54	-36.30
	4	3	-	7	1	14	89	45	5.55	-36.80
SMP DPD	2	4	-	7	1	30	65	50	5.55	-37.20
	3	4	-	7	1	31	99	63	5.57	-37.80
	4	4	-	7	1	32	143	77	5.57	-37.80
MP DPD	11	4	0	-	-	24	255	112	5.47	-38.20
GMP DPD	11	4	9	-	-	42	393	184	5.44	-38.40

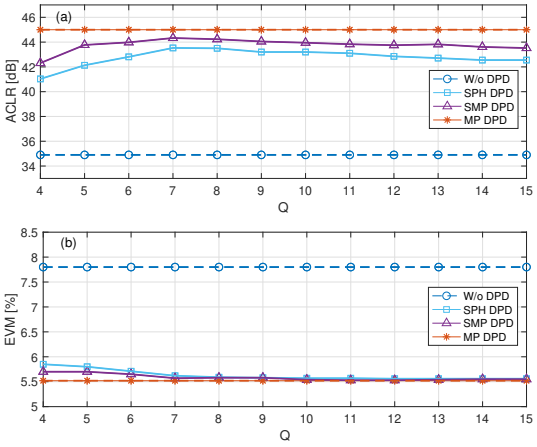


Fig. 7. Measured ACLR and EVM performance in Experiment 1 at PA output power of +27 dBm, as a function of the number of LUT control points in the SPH and SMP models, Q, while adopting $P_{\mathrm{SP}} = 3$, $M_{\mathrm{SPH}} = 3$, $M_{\mathrm{SMP}} = 4$, $P_{\mathrm{MP}} = 11$, and $M_{\mathrm{MP}} = 4$.

behave very similarly. Similar observations follow also from Fig. 7, showing again the EVM and ACLR metrics but this time at fixed PA output power of +27 dBm while then varying the number of LUT control points in the proposed SPH and SMP models. From this figure we can also observe that the LUT based DPD performance is optimized with approximately $Q = 7$ or $Q = 8$ control points, in this example, while in general it is likely that the optimization of the value of Q is to be done separately for different PA types.

Finally, Table III then collects and summarizes the obtained DPD results in Experiment 1 while also showing the DPD main path processing complexities. Here also other spline interpolation orders P_{SP} are considered and shown. We can conclude that the proposed spline-based DPD models offer a favorable performance-complexity trade-off compared to the reference MP DPD approach.

C. Experiment 2: 5G NR Band 78 Small-Cell PA

The second experiment includes the Skyworks SKY66293-21 PA module, illustrated in Fig. 4(c), which is a low-to-medium output power amplification unit oriented to be used either in small-cell base-stations or in large antenna array transmitters. The PA module is specifically designed to operate in the NR Band n78 (3300-3800 MHz), having a gain of 34 dB, and a 1-dB compression point of +31.5 dBm. Similar 5G NR downlink signal corresponding to the 100 MHz channel bandwidth scenario, as in the Experiment 1, is adopted, while the considered RF center-frequency is 3.65 GHz. The test signal is again transmitted via the RF TX port of the VST directly to the PA module, while the considered PA output power is +24 dBm, corresponding to the maximum transmit power of a Local Area BS according to the NR regulations [23]. Again, five ILA learning iterations are adopted while the signal length within each ILA iteration is 100,000 samples. The VST observation receiver runs at 491.52 MHz ($4\times$ oversampling).

Fig. 8 and Table IV illustrate and summarize the obtained linearization results for the proposed and the reference DPD methods. Again, also comparative complexity numbers are stated in Table IV. As stated in [23], a 5G NR Local Area BS can operate within an absolute basic limit of -32 dBm/MHz in the adjacent channel region, assuming the considered PA output power of +24 dBm. As shown in Fig. 8 and Table IV, the SPH, SMP, MP, and GMP DPD satisfy this limit, indicating successful linearization. Table IV shows again that remarkable complexity reductions are obtained through the proposed spline-based DPD approaches, compared to the reference MP or GMP DPDs, while all provide a very similar linearization performance.

D. Experiment 3: FR-2 Environment and 28 GHz Active Array

In order to further demonstrate the applicability of the proposed spline-based DPD concepts, the third and final experiment focuses on timely 5G NR mmWave/FR-2 deployments [23] with active antenna arrays. Unwanted emission modeling and DPD-based linearization of active arrays with large numbers of PA units is, generally, an active research field, with good examples of recent papers being, e.g., [5], [43]–[48]. Below we first describe shortly the FR-2 measurement setup, and then present the actual linearization results.

1) FR-2 Measurement Setup: The overall mmWave/FR-2 measurement setup is depicted in Fig. 9, incorporating an Anokiwave AWMF-0129 active antenna array, together with

TABLE IV
SUMMARY OF DPD MAIN PATH PROCESSING COMPLEXITY AND LINEARIZATION PERFORMANCE IN EXPERIMENT 2, PA OUTPUT POWER IS +24 dBm

	Running complexity								Model performance	
	P	**M**	**G**	**Q**	$\Delta_{z,x}$	**# of coefficients**	**FLOPs/sample**	**Mul./sample**	**EVM (%)**	**Max. dBm/MHz**
No DPD	-	-	-	-	-	-	-	-	8.64	-18.20
SPH DPD	2	4	-	7	1	13	63	32	5.70	-31.40
	3	4	-	7	1	14	77	40	5.57	-33.20
SMP DPD	2	5	-	7	1	37	73	56	5.60	-32.90
	3	5	-	7	1	38	111	75	5.55	-33.10
MP DPD	11	5	0	-	-	30	303	136	5.54	-33.20
MP DPD	11	5	9	-	-	48	441	208	5.49	-39.80

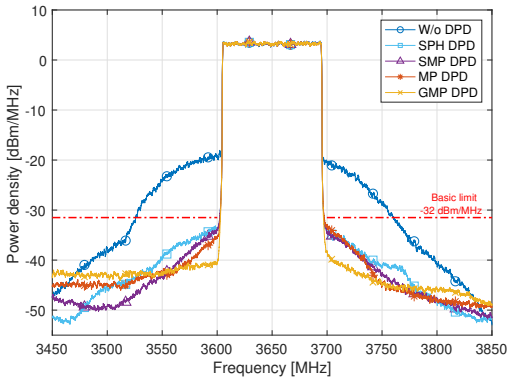


Fig. 8. Example illustration of linearization results in Experiment 2 (NR small-cell PA measured at 3.65 GHz), with 100 MHz channel bandwidth and PA output power of +24 dBm, while adopting $P_{\mathrm{SP}} = 3$, $Q_{\mathrm{SPH}} = Q_{\mathrm{SMP}} = 7$, $M_{\mathrm{SPH}} = 4$, $M_{\mathrm{SMP}} = 5$, $P_{\mathrm{MP}} = P_{\mathrm{GMP}} = 11$, $M_{\mathrm{MP}} = M_{\mathrm{GMP}} = 5$, and $G = 9$.

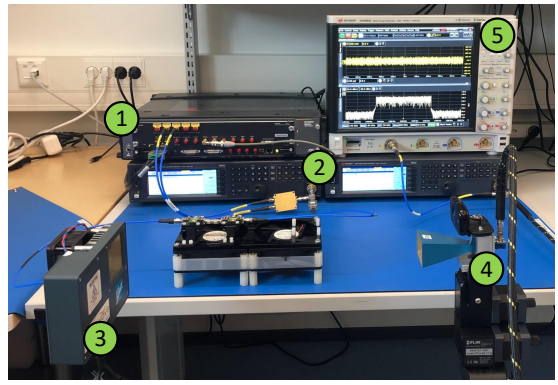


Fig. 9. RF measurement setup in Experiment 3 including the Keysight M8190 arbitrary waveform generator (1), Keysight N5183B-MXG LO signal generators (2), Anokiwave AWMF-0129 active antenna array (3) working at 28 GHz center frequency (NR Band n257), horn antenna as receiver (4), and Keysight DSOS804A digitizer (5).

other relevant instruments and equipment for signal generation and analysis, facilitating measurements at 28 GHz center-frequency with up to 3 GHz of instantaneous bandwidth. On the transmit chain side, the setup consists of a Keysight M8190 arbitrary waveform generator that is used to generate directly the I/Q samples of a wideband modulated IF signal centered at 3.5 GHz. The signal is then upconverted to the 28 GHz carrier frequency by utilizing the Keysight N5183B-MXG that generates the corresponding local oscillator signal running at 24.5 GHz, together with external mixers and filters. The modulated RF waveform is then pre-amplified by means of two Analog Devices' driver PAs, HMC499LC4 and HMC943ALP5DE, with 17 dB and 23 dB gain, respectively, such that the integrated PAs of the Anokiwave AWMF-0129 active antenna array are driven towards saturation. The 1 dB compression point of the active array, when measured in terms of the effective isotropic radiated power (EIRP) and 5G NR compliant OFDM waveform, is approximately +42 dBm.

The transmit signal propagates over-the-air (OTA) and is captured by a horn antenna at the observation receiver, such that the receiving antenna system is well aligned with the main transmit beam. At the receiver side, another Keysight N5183B-MXG and a mixing stage are used to downconvert the signal back to IF. Then, the Keysight DSOS804A oscilloscope is utilized as the actual digitizer, including also built-in filtering, and the signal is taken to baseband and processed in a host PC, where the DPD learning and predistortion are executed. The OTA measurement system is basically following the measurement procedures described in [23], [43], [49], specifically the measurement option utilizing the beam-based directions. In these measurements, the observation receiver provides I/Q samples at 7× oversampled rate.

2) Active Array Linearization: Linearization of active phased-array transmitters is generally a challenging task, since a single DPD unit must linearize a bank of mutually different PAs. There are multiple ways of acquiring the observation signal for DPD parameter learning, as discussed e.g. in [5], [43]–[47]. In this work, we assume and adopt the so-called combined observation signal approach and utilize specifically the OTA-combined received signal for DPD parameter learning [5], [44], [47], while otherwise following exactly the same learning algorithms as in the Experiments 1 and 2.

In the DPD measurements, we adopt 5G NR FR-2 OFDM signal with SCS of 60 kHz and 7 dB PAPR (0.01% CCDF point), and consider active PRB counts of 132 and 264, mapping to 100 MHz and 200 MHz channel bandwidths, respectively [23]. In this case, 5 ILA iterations are adopted,

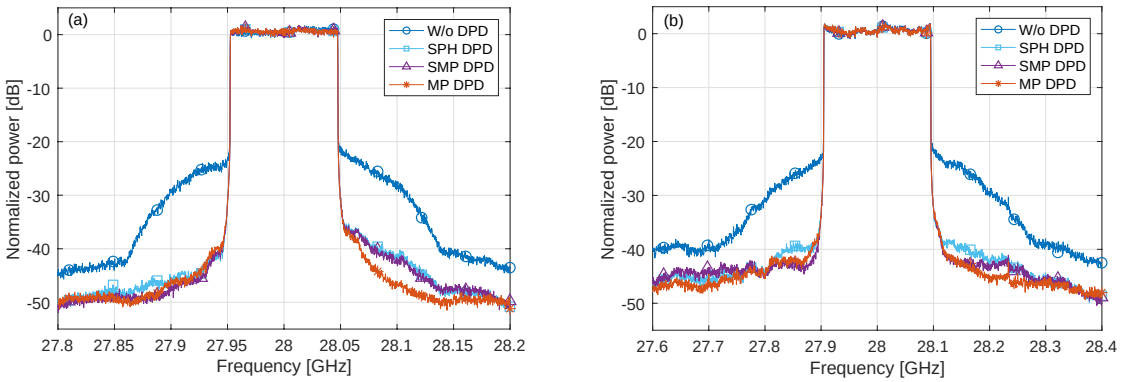


Fig. 10. Illustration of OTA linearization of the Anokiwave AWMF-0129 active antenna array, when (a) NR 100 MHz and (b) NR 200 MHz transmit signals are applied, measured at EIRP of +42.5 dBm The SPH and SMP DPD spline order is $P_{\mathrm{SP}} = 3$, while $Q_{\mathrm{SPH}} = Q_{\mathrm{SMP}} = 7$, $M_{\mathrm{SPH}} = 3$, and $M_{\mathrm{SMP}} = 4$. The MP DPD order is $P_{\mathrm{MP}} = 11$ while $M_{\mathrm{MP}} = 4$.

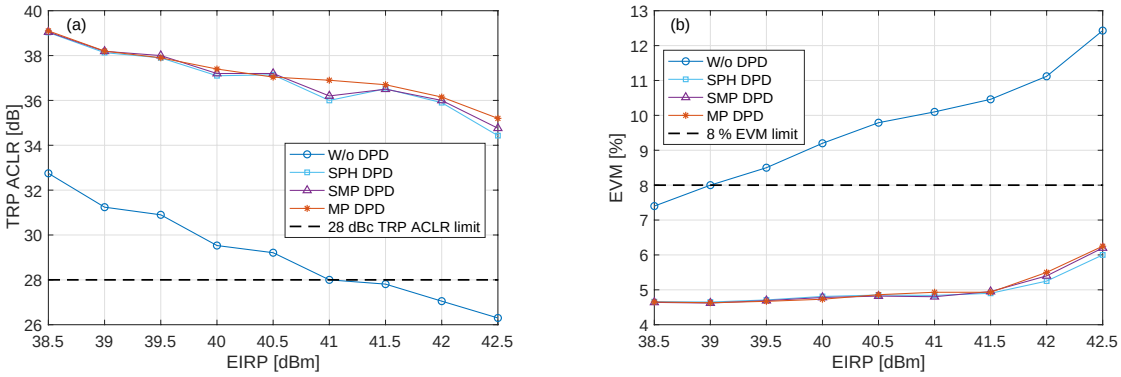


Fig. 11. OTA linearization performance of the Anokiwave AWMF-0129 active antenna array, with NR 100 MHz OFDM waveform, as a function of the EIRP of the proposed DPD models in terms of a) TRP ACLR, and b) EVM.

TABLE V
SUMMARY OF LINEARIZATION PERFORMANCE OF THE ANOKIWAVE AWMF-0129 ACTIVE ANTENNA ARRAY, WITH 100 MHZ AND 200 MHZ 5G NR CHANNEL BANDWIDTHS, MEASURED AT +42.5 dBm EIRP

	DPD running complexity						DPD perf., 100 MHz		DPD perf., 200 MHz	
	P	**M**	**Q**	$\Delta_{z,x}$	**FLOPs/sample**	**Mul./sample**	**EVM (%)**	**TRP ACLR (dB)**	**EVM (%)**	**TRP ACLR (dB)**
No DPD	-	-	-	-	0	0	12.10	26.10	12.43	26.30
SPH DPD	3	3	7	1	69	36	6.20	34.40	6.25	34.10
SMP DPD	3	4	7	1	99	63	6.15	34.80	6.20	34.40
MP DPD	11	4	-	-	255	112	6.00	35.20	6.13	35.00

each containing 50,000 samples. Example OTA linearization results are illustrated in Fig. 10, measured at an EIRP of +42.5 dBm, where the received spectra with the proposed SPH, SMP and the reference MP DPD are shown, while the no-DPD case is also shown for comparison. The parametrization of the SPH and SMP DPD is $P_{\mathrm{SP}} = 3$ and $M_{\mathrm{SPH}} = 3$, and $M_{\mathrm{SMP}} = 4$, while MP DPD is configured with $P_{\mathrm{MP}} = 11$ and $M_{\mathrm{MP}} = 4$. As mentioned already in the introduction, the OTA ACLR requirements at FR-2 are quite clearly relaxed, compared to the classical 45 dB at FR-1, with 28 dB defined as the TRP-based ACLR limit in the current NR Release-15 specifications [23]. Additionally, 64-QAM is currently the highest supported modulation scheme at FR-2, heaving 8% as the required EVM.

In both channel bandwidth cases, considered in Fig. 10, the initial EVM and TRP ACLR metrics are around 12.5% and 26 dB, respectively, when measuring at EIRP of +42.5 dBm and when no DPD is applied. Hence, linearization is indeed required if the same output power is to be maintained, while Fig. 10 demonstrates that all considered DPD methods can successfully linearize the active array. Table V shows the exact measured numerical TRP ACLR and EVM values, indicating

good amounts of linearization gain and that the EVM and TRP ACLR requirements can be successfully met. It is also noted that the initial TRP ACLR of approximately 26 dB corresponds already to a very nonlinear starting point.

Finally, Fig. 11 features a power sweep performed with the antenna array, with NR 100 MHz OFDM waveform, illustrating the TRP ACLR and EVM as a function of the EIRP with and without DPD. It can be clearly observed that in this particular experiment, when no DPD is applied, it is the EVM metric that is limiting the maximum achievable EIRP such that both TRP ACLR and EVM requirements are still fulfilled. Specifically, without DPD processing, this limits the maximum EIRP to approximately +39 dBm, while when DPD processing is applied, both requirements are fulfilled at least up to the considered maximum EIRP of +42.5 dBm – and clearly also beyond. In this particular linearization experiment, the SPH DPD provides an excellent linearization performance, despite its very low computational complexity.

VI. Conclusions

In this paper, novel complex spline-interpolated LUT concepts and corresponding DPD methods with gradient-adaptive learning rules were proposed for power amplifier linearization. A vast amount of different measurement-based experiments were provided, covering successful linearization of different PA samples at sub-6 GHz bands. Additionally, a 28 GHz state-of-the-art active antenna array was successfully linearized. The measured linearization performance results, together with the provided explicit complexity analyses, show that the proposed spline-interpolated DPD concepts can provide very appealing complexity-performance trade-offs, compared to, e.g., ordinary canonical MP DPD. Specifically, the SMP DPD was shown to provide in all measurement examples linearization performance very close to that of ordinary MP DPD, while having substantially lower main path and DPD learning complexity. Additionally, the SPH DPD offers further reduction in the main path processing complexity, while demonstrating a linearization performance that is fairly close to the other DPD systems, particularly in the timely 28 GHz active array linearization experiment. Our future work topics contain extending the current work to direct learning architecture (DLA) context and to digital MIMO transmitters with multi-dimensional DPD basis functions.

References

[1] E. Dahlman, S. Parkvall, and J. Skold, *5G NR: The Next Generation Wireless Access Technology*, 1st ed. Academic Press, 2020.

[2] F. M. Ghannouchi, and O. Hammi, "Behavioral modeling and predistortion", *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52-64, Dec. 2009.

[3] 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, Mar. 2013.

[4] S. Afsardoost, T. Eriksson, and C. Fager. "Digital Predistortion Using a Vector-Switched Model", *IEEE Trans. Microw. Theory Tech.*, vol 60, no. 4, pp. 1166-1174, Apr. 2012.

[5] C. Fager *et al*., "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.

[6] O. Hammi, F. M. Ghannouchi and B. Vassilakis, "A Compact Envelope-Memory Polynomial for RF Transmitters Modeling With Application to Baseband and RF-Digital Predistortion", *IEEE Microw. Wireless Compon. Lett.*, vol. 18, no. 5, pp. 359-361, May 2008.

[7] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path", *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 280-293, Jan. 2018.

[8] Y. Ma, Y. Yamao, Y. Akaiwa, and C. Yu, "FPGA Implementation of Adaptive Digital Predistorter With Fast Convergence Rate and Low Complexity for Multi-Channel Transmitters", *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 11, pp. 3961-3973, Nov. 2013.

[9] J. Kim, and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory", *Electron. Lett.*, vol. 37, no. 23, pp. 1, Nov. 2001.

[10] A. Zhu, "Digital Predistortion and Its Combination with Crest Factor Reduction", in *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*, F.-L. Lue, Ed. Cambridge: Cambridge University Press, 2011, ch. 9, pp. 244-279.

[11] A. S. Tehrani et al., "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models", *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 6, pp. 1510-1520, Jun. 2010.

[12] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers", *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852-3860, Oct. 2006.

[13] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-band complexity-reduced generalized-memory-polynomial power-amplifier digital predistortion", *IEEE Trans. Microw. Theory Tech.*, vol. 64, no. 6, pp. 1763-1774, Jun. 2016.

[14] A. Zhu and T. J. Brazil, "An Overview of Volterra Series Based Behavioral Modeling of RF/Microwave Power Amplifiers", in *2006 IEEE Annual Wireless and Microwave Technology Conference*, Dec. 2006, pp. 1-5.

[15] ——, "Behavioral modeling of RF power amplifiers based on pruned volterra series", *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 12, pp. 563-565, Dec. 2004.

[16] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Melbourne, FL, USA: Krieger Publishing Co., Inc., 2006.

[17] C. Yu, L. Guan, E. Zhu and A. Zhu, "Band-Limited Volterra Series-Based Digital Predistortion for Wideband RF Power Amplifiers", *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 12, pp. 4198-4208, Dec. 2012.

[18] N. Safari, N. Holte and T. Roste, "Digital Predistortion of Power Amplifiers Based on Spline Approximations of the Amplifier Characteristics", in *2007 IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 2075-2080.

[19] N. Safari, P. Fedorenko, J. S. Kenney and T. Roste, "Spline-Based Model for Digital Predistortion of Wide-Band Signals for High Power Amplifier Linearization", in *2007 IEEE/MTT-S International Microwave Symposium*, Jun. 2007, pp. 1441-1444.

[20] 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*, vol. 67, no. 6, Jun. 2020.

[21] C. Cheang, P. Mak and R. P. Martins, "A Hardware-Efficient Feedback Polynomial Topology for DPD Linearization of Power Amplifiers: Theory and FPGA Validation", *IEEE Trans. Circuits Syst. I*, vol. 65, no. 9, pp. 2889-2902, 2018.

[22] F. M. Barradas, T. R. Cunha, P. M. Lavrador and J. C. Pedro, "Polynomials and LUTs in PA Behavioral Modeling: A Fair Theoretical Comparison", *IEEE Trans. Microw. Theory Tech.*, vol. 62, no. 12, pp. 3274-3285, Nov. 2014.

[23] 3GPP Tech. Spec. 38.104, "NR; Base Station (BS) radio transmission and reception", v15.4.0 (Release 15), Dec. 2018.

[24] K.-F. Liang, J.-H. Chen, and Y.-J. E. Chen, "A quadratic-interpolated LUT-based digital predistortion technique for cellular power amplifiers", *IEEE Trans. Circuits Syst. II*, vol. 61, no. 3, pp. 133-137, Mar. 2014.

[25] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation", *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980-987, Nov. 2017.

[26] P. Jardin and G. Baudoin, "Filter lookup table method for power amplifier linearization", *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1076-1087, May. 2007.

[27] X. Wu, N. Zheng, X. Yang, J. Shi and H. Chen, "A spline-based Hammerstein predistortion for 3G power amplifiers with hard nonlinearities", in *2010 2nd Int. Conf. Future Comp. Commun.*, vol. 3, May. 2010.

[28] O. Hammi, F. M. Ghannouchi, S. Boumaiza and B. Vassilakis, "A Data-Based Nested LUT Model for RF Power Amplifiers Exhibiting Memory Effects", *IEEE Microw. Wireless Compon. Lett.*, vol. 17, no. 10, pp. 712-714, Oct. 2007.

[29] H. Zhi-yong *et al*., "An improved look-up table predistortion technique for HPA with memory effects in OFDM systems", *IEEE Trans. Broadcasting*, vol. 52, no. 1, pp. 87-91, Feb. 2006.

[30] 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 Tech.*, vol. 66, no. 12, 2018.

[31] Z. Wang *et al*., "Low Computational Complexity Digital Predistortion Based on Direct Learning With Covariance Matrix", *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 11, pp. 4274-4284, May 2018.

[32] C. De Boor, *A Practical Guide to Splines*, Springer, New York, 1978.

[33] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Nonlinear spline adaptive filtering", *Signal Process.*, vol. 93, no. 4, pp. 772-783, Apr. 2013.

[34] ——, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties", *Signal Process.*, vol. 100, pp. 112-123, Jul. 2014.

[35] ——, "Novel cascade spline architectures for the identification of nonlinear systems", *IEEE Trans. Circuits Syst. I*, vol. 62, no. 7, pp. 1825-1835, Jul. 2015.

[36] P. Pascual Campo, D. Korpi, L. Anttila, and M. Valkama, "Nonlinear Digital Cancellation in Full-Duplex Devices using Spline-Based Hammerstein Model", in *Proc. IEEE GLOBECOM 2018*, Dec. 2018.

[37] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 2001.

[38] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory", *IEEE Proceedings F - Commun., Radar and Signal Process.*, vol. 130, no. 1, pp. 11-16, Feb. 1983.

[39] M. Frerking, *Digital Signal. Processing in Communications Systems*, Springer, 1994.

[40] R. Raich, Hua Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design", *IEEE Trans. Veh. Technol.*, vol. 53, no. 5, pp. 1468-1479, Sep. 2004.

[41] R. Dallinger, H. Ruotsalainen, R. Wichman and M. Rupp, "Adaptive predistortion techniques based on orthogonal polynomials", in *44th Asilomar Conf. Signals, Syst., and Computers*, 2010, pp. 1945-1950.

[42] R. J. Shilling and S. L. Harris, *Fundamentals of Digital Signal Processing Using MATLAB*, Cengage Learning, 2010.

[43] A. Brihuega, M. Abdelaziz, L. Anttila, M. Turunen, M. Allén, 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, 2020.

[44] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson and M. Valkama, "Digital Predistortion for Hybrid MIMO Transmitters", *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445-454, Jun. 2018.

[45] A. Brihuega, L. Anttila, M. Abdelaziz and M. Valkama, "Digital Predistortion in Large-Array Digital Beamforming Transmitters", in *52nd Asilomar Conf. Signals, Syst., Comp.*, Oct. 2018, pp. 611-618.

[46] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi and Z. Feng, "Beam-Oriented Digital Predistortion for 5G Massive MIMO Hybrid Beamforming Transmitters", *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 7, pp. 3419-3432, Jul. 2018.

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

[48] 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.

[49] 3GPP Tech. Spec. 38.141-2, "NR; Base Station (BS) conformance testing, Part 2", v15.1.0 (Release 15), Mar. 2019.

Pablo Pascual Campo received his B.Sc. and M.Sc. degrees in Telecommunications and Electrical Engineering in 2012 and 2014, respectively, from Universidad Politécnica de Madrid, Madrid, Spain. He is currently pursuing his D.Sc. degree at Tampere University, Department of Electrical Engineering, Tampere, Finland. His research interests include digital predistortion, full-duplex systems and applications, and signal processing for wireless communications at the mmWave bands.

Alberto Brihuega (S'18) received the B.Sc. and M.Sc. degrees in Telecommunications Engineering from Universidad Politécnica de Madrid, Spain, in 2015 and 2017, respectively. He is currently working towards the Ph.D. degree with Tampere University, Finland, where he is a researcher with the Department of Electrical Engineering. His research interests include statistical and adaptive digital signal processing for compensation of hardware impairments in large-array antenna transceivers.

Lauri Anttila received the D.Sc. (Tech.) degree (with distinction) in 2011 from Tampere University of Technology (TUT), Finland. Since 2016, he has been a University Researcher at the Department of Electrical Engineering, Tampere University (formerly TUT). In 2016-2017, he was a Visiting Research Fellow at Aalto University, Finland. His research interests are in radio communications and signal processing, with a focus on the radio implementation challenges in systems such as 5G, full-duplex radio, and large-scale antenna systems.

Matias Turunen is currently pursuing the M.Sc. degree in electrical engineering at Tampere University (TAU), Tampere, Finland, while also working as a Research Assistant with the Department of Electrical Engineering at TAU. His current research interests include in-band full-duplex radios with an emphasis on analog RF cancellation, OFDM radar, and 5G New Radio systems.

Dani Korpi received D.Sc. degree (Hons.) in electrical engineering from Tampere University of Technology (TUT), Finland, in 2017. Currently, he is a Senior Specialist with Nokia Bell Labs in Espoo, Finland. He received the best dissertation award of the year in TUT, as well as the Finnish Technical Sector's Award for the best doctoral dissertation of 2017. His research interests include inband full-duplex radios, machine learning for wireless communications, and beyond 5G radio systems.

Markus Allén received the B.Sc., M.Sc. and D.Sc. degrees in communications engineering from Tampere University of Technology, Finland, in 2008, 2010 and 2015, respectively. He is currently with the Department of Electrical Engineering at Tampere University, Finland, as a University Instructor. His current research interests include software-defined radios, 5G-related RF measurements and digital signal processing for radio transceiver linearization.

Mikko Valkama received the D.Sc. (Tech.) degree (with honors) from Tampere University of Technology (TUT), Finland, in 2001. In 2003, he was a visiting post-doc research fellow with SDSU, San Diego, CA. Currently, he is a Full Professor and Department Head of Electrical Engineering at the newly formed Tampere University (TAU), Finland. His research interests include radio communications, radio localization, and radio-based sensing, with emphasis on 5G and beyond mobile radio networks.

PUBLICATION 2

Closed-loop sign algorithms for low-complexity digital predistortion: methods and performance

P. Pascual Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allén, Y. Guo and M. Valkama

IEEE Transactions on Microwave Theory and Techniques 69.1 (2021), 1048–1062
DOI: 10.1109/TMTT.2020.3038316

Publication reprinted with the permission of the copyright holders

Closed-Loop Sign Algorithms for Low-Complexity Digital Predistortion: Methods and Performance

Pablo Pascual Campo, *Student Member, IEEE,* Vesa Lampu, Lauri Anttila, *Member, IEEE,* Alberto Brihuega, *Student Member, IEEE,* Markus Allén, Yan Guo, and Mikko Valkama, *Senior Member, IEEE*

***Abstract*—In this article, we study digital predistortion (DPD) based linearization with specific focus on millimeter wave (mmW) active antenna arrays. Due to the very large channel bandwidths and beam-dependence of nonlinear distortion in such systems, we present a closed-loop DPD learning architecture, look-up table (LUT) based memory DPD models, and low-complexity sign-based estimation algorithms, such that even continuous DPD learning could be technically feasible. To this end, three different learning algorithms – Sign, Signed Regressor, and Sign-sign – are formulated for the LUT-based DPD models, such that the potential rank deficiencies, experienced in earlier methods, are avoided, while facilitating greatly reduced learning complexity. The injection-based LUT DPD structure is also shown to allow for low numbers and reduced dynamic range of the involved LUT entries. Extensive RF measurements utilizing a state-of-the-art mmW active antenna array system at 28 GHz are carried out and reported to validate the methods, incorporating very wide channel bandwidths of 400 MHz and 800 MHz while pushing the array close to saturation. Additionally, the processing and learning complexities of the considered techniques are analyzed, which together with the measured linearization performance figures allow to assess the complexity-performance trade-offs of the proposed solutions. Overall, the results show that efficient mmW array linearization can be obtained through the proposed methods at very low complexity.**

***Index Terms*—ACLR, active array transmitters, closed-loop systems, digital predistortion, EVM, lookup table, millimeter-wave frequencies, nonlinear distortion, over-the-air, sign algorithms, signed regressor.**

I. Introduction

THE adoption of modern, spectrally efficient waveforms with high peak-to-average power ratio (PAPR), most notably OFDM, complicates operating power amplifiers (PAs) close to saturation [1]. To ensure a good power efficiency, while at the same time controlling the transmitted signal quality, digital predistortion (DPD) based linearization is a well-known and widely-applied approach, see, e.g., [1] and [2] and the references therein. DPD aims at suppressing the unwanted out-of-band (OOB) emissions and passband nonlinear distortion steaming from the PAs by applying an appropriate nonlinear transformation to the digital transmit waveform. Especially when combined with PAPR reduction methods [2], the DPD system can largely improve the transmitter power efficiency, while maintaining the passband signal quality and OOB emissions within specified limits [3], [4].

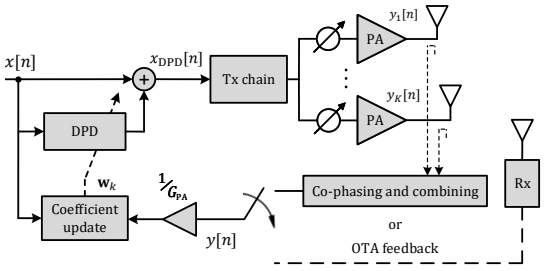


Fig. 1. Illustration of the injection-based DPD scheme with closed-loop parameter learning for linearizing an active phased-array transmitter, with K antennas and PA units. Observation path builds on co-phasing and combining the PA output signals, or alternatively on OTA feedback.

One modern and timely DPD use case is the linearization of active antenna array based base-stations of the emerging 5G New Radio (NR) networks at millimeter-wave (mmW) bands – referred to as frequency range 2, FR-2 – with good examples of recent papers being [5]–[13]. In such DPD systems, stemming from the load modulation phenomenon, the effective nonlinear distortion has been observed to be clearly beam-dependent [9], and thus fast DPD adaptation is required. This issue, together with the very wide channel bandwidths [14], and thus DPD processing rates, calls for low-complexity DPD systems and parameter learning algorithms. Such methods are currently under intensive research and form also the topic of this paper.

In the existing literature, various DPD architectures and PA modeling methods have been widely studied, with the memory polynomial (MP) [1], [4], [15] and the generalized memory polynomial (GMP) [1], [4], [16] being some of the most common approaches. Both of these techniques can be interpreted to be subsets of the Volterra series [1], [17], [18]. While these approaches typically provide an accurate and reliable DPD linearization performance, they often involve a relatively high processing complexity, which can pose a challenge for real-time implementations.

The literature on low-complexity DPD methods and the

Manuscript received July 21, 2020; revised October 20, 2020; accepted October 30, 2020. This article is an extended version from the 2020 IEEE MTT-S International Microwave Symposium (IMS-2020) [28]. (Corresponding author: Pablo Pascual Campo)

This work was financially supported by the Academy of Finland under the projects 301820, 323461, 332361 and 319994.

P. Pascual Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allén, and M. Valkama are with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland (e-mail: pablo.pascualcampo@tuni.fi.).

Y. Guo is with the Wireless Terminal Algorithm Development Department, HiSilicon Technologies Co. Ltd, Shanghai, China.

associated learning algorithms is, on the other hand, somewhat more scarce. Techniques towards this direction are, for instance, [7], [18]–[25]. In [7], a reduced complexity approach which utilizes the combined PA output signals together with a computationally efficient closed-loop learning equation to minimize the distortion in the main beam direction, is investigated. In [20], in a more traditional single-antenna DPD context, the use of 1-bit observations in closed-loop learning is considered, in combination with a sign-based Gauss-Newton (GN) learning algorithm. In [21], a GN signed regressor algorithm (SRA) is formulated for real-valued feedback signals. The signed regressor matrix is, however, rank deficient, and thus an additional Walsh-Hadamard transformation is applied to make it invertible, further increasing the computational complexity. In [22], a look-up table (LUT) based MP DPD with a sample-adaptive least mean squares (LMS) SRA is proposed. However, in this work each LUT in the MP structure is updated independently, making the solution sub-optimal. In [23], direct least squares (LS) and GN adaptations for linearly interpolated LUT-based Volterra models are proposed in indirect learning architecture (ILA) and closed-loop context, respectively. In [24], [26], cascaded Hammerstein structures with polynomial and spline nonlinearities were proposed. Cascaded structures typically have less free parameters, making them appealing when low-complexity solutions are pursued. However, the models were based on the ILA in combination with LS-based learning algorithms, which complicates adaptive estimation and tracking.

In this article, contrary to the earlier closed-loop works in [20]–[23], we adopt the so-called injection-based DPD structure [7], [27], illustrated in Fig. 1 in the context of mmW active arrays. To this end, building on our early work in [28], we formulate various signed learning methods – Sign, Signed Regressor, and Sign-sign – based on the GN, self-orthogonalization (SO), and block-LMS (BLMS) learning rules. Such sign algorithms allow for a large complexity reduction in the DPD learning, since the needed number of multiplications is largely reduced compared to the reference methods. Additionally, we adopt a LUT-based memory DPD model. LUT-based structures are generally simpler than polynomial-type ones used in the reference works [7], [20], [21], allowing large reductions in terms of the processing and learning complexities. Furthermore, adopting the injection-based DPD allows to significantly reduce the LUT sizes, such that 32 or even 16 entries are enough for efficient linearization, without interpolation. Additionally, the use of non-interpolated LUTs avoids the rank deficiencies in the SRA and Sign-sign algorithms and thus the additional matrix transformation, which were experienced in [21]. Due to their low complexity and closed-loop nature, the developed solutions allow for fast real-time adaptation, and thus potentially on-chip implementations and continuous learning. We also show that the injection-based DPD formulation allows for dynamic range reduction in the LUT control points, and thus facilitates efficient fixed-point implementations with relatively low number of bits.

Extensive RF measurement results at 28 GHz (5G NR band n257 [14]), utilizing a state-of-the-art 64-element active antenna array and 5G NR like OFDM waveforms, are reported and analyzed, incorporating standard-compliant channel bandwidth of 400 MHz while also pushing the performance boundaries further up to 800 MHz. The obtained linearization results, together with the provided detailed complexity analysis, show that the proposed methods provide very favorable complexity-performance trade-offs, while meeting the 3GPP 5G NR [14] OOB emission and passband transmit signal quality requirements at FR-2 in all tested scenarios, even in the ambitious 800 MHz channel bandwidth case. Overall, the results show that efficient mmW array linearization can be obtained through the proposed methods.

In short, the novelty and contributions of the article can be summarized as follows:

- Injection-based memory polynomial LUT DPD system is proposed, shown to significantly reduce the LUT entry sizes to achieve efficient linearization. Additionally, the injection-based scheme is also shown to allow for dynamic range reduction in the LUT control points, thus facilitating efficient fixed-point implementations;
- Various sign-based low-complexity closed-loop learning algorithms are formulated in the context of injection-based MP LUT DPD system;
- Extensive computational complexity analysis of the different signed learning rules is provided and also compared to the corresponding unsigned algorithms;
- Very extensive 28 GHz active array linearization measurements are provided and analyzed, incorporating channel bandwidth up to 800 MHz;

It is finally noted that even though our primary applications are in the mmW active array transmitters, the proposed techniques are applicable to any single-input single-output DPD system, where the PA output is commonly observed directly through a directional coupler and an observation receiver.

The rest of the paper is organized as follows. Section II first presents the proposed injection-based MP-LUT closed-loop DPD system, together with the unsigned GN, SO, and BLMS learning principles. Additionally, the dynamic range reduction through the injection-based DPD approach is addressed, while the different options for arranging the DPD feedback signal in mmW active arrays are also shortly discussed. Section III then describes the different sign-based learning algorithms. Section IV presents a detailed complexity analysis of the considered unsigned and signed learning rules. Section V presents an extensive set of RF measurements at 28 GHz which test and validate the proposed approaches. Finally, Section VI concludes the paper.

Notation Used in This Article

In this paper, matrices are represented by capital boldface letters, i.e., $\mathbf{\Sigma} \in \mathbb{C}^{M\times N}$. Ordinary transpose, Hermitian transpose, and complex conjugation are denoted by $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. By default, vectors are complex-valued column vectors, presented with lowercase boldface letters, i.e., $\mathbf{v} \in \mathbb{C}^{M\times 1} = [v_1 \; v_2 \; \cdots \; v_M]^T$. Additionally, the absolute value and floor operators are represented as $|\cdot|$ and $\lfloor\cdot\rfloor$, respectively.

$$\mathbf{\Phi} = \begin{bmatrix} x[n]\boldsymbol{\xi}_n^T(|x[n]|,k) & x[n-1]\boldsymbol{\xi}_{n-1}^T(|x[n-1]|,k) & \cdots & x[n-M+1]\boldsymbol{\xi}_{n-M+1}^T(|x[n-M+1]|,k) \\ x[n+1]\boldsymbol{\xi}_{n+1}^T(|x[n+1]|,k) & x[n]\boldsymbol{\xi}_n^T(|x[n]|,k) & \cdots & x[n-M+2]\boldsymbol{\xi}_{n-M+2}^T(|x[n-M+2]|,k) \\ \vdots & \vdots & \ddots & \vdots \\ x[n+N-1]\boldsymbol{\xi}_{n+N-1}^T(|x[n+N-1]|,k) & x[n+N-2]\boldsymbol{\xi}_{n+N-2}^T(|x[n+N-2]|,k) & \cdots & x[n+N-M]\boldsymbol{\xi}_{n+N-M}^T(|x[n+N-M]|,k) \end{bmatrix} \tag{1}$$

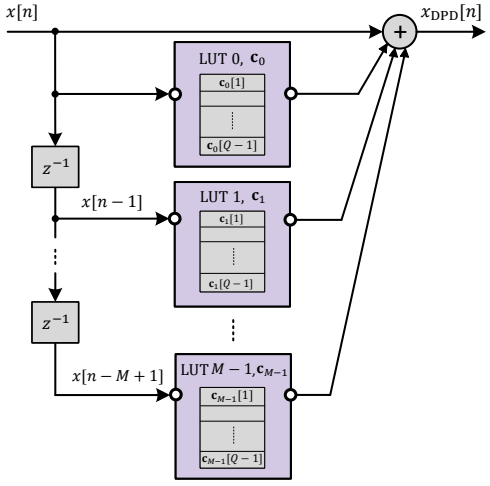


Fig. 2. The input-output relation of the proposed MP-LUT DPD model, in combination with the injection-based scheme.

II. Closed-Loop DPD System

In this work, we adopt and formulate the MP DPD model, where the high-order polynomial functions are replaced with Q entry-sized LUTs [23]. This model is adopted due to its inherent low processing complexity, [22], [29]. Additionally, the system builds on a closed-loop learning architecture, where the DPD coefficients are directly adapted using the input signal $x[n]$ and the observed signal $y[n]$ [30], following the basic notations shown in Fig. 1.

A. Injection Based MP-LUT DPD

Formally, the input-output relation of an ordinary MP can be formulated as a function of its polynomial order P and memory-depth M, as

$$x_{\mathrm{DPD}}[n] = \sum_{m=0}^{M-1} \sum_{\substack{p=0 \\ p\,\mathrm{odd}}}^{P} \alpha_{m,p} x[n-m]|x[n-m]|^{(p-1)}, \tag{2}$$

where $\alpha_{m,p}$ is the corresponding PA model coefficient. In order to substitute the polynomials with LUTs, (2) can be rewritten as

$$x_{\mathrm{DPD}}[n] = \sum_{m=0}^{M-1} x[n-m] G_m(|x[n-m]|). \tag{3}$$

Herein, $G_m(|x[n-m]|),\ m = 0, 1, \cdots, M-1$, refer to the complex LUT gains, weighting the input samples in each memory branch, denoted here by the parameter m. This complex LUT gain can be defined and expressed as

$$G_m(|x[n-m]|) = \boldsymbol{\xi}_{n-m}^T(|x[n-m]|,k)\mathbf{c}_m, \tag{4}$$

where $\mathbf{c}_m \in \mathbb{C}^{Q\times 1},\ m = 0, 1, \cdots, M-1$, are the M corresponding Q-sized LUTs, while the vector $\boldsymbol{\xi}_n(|x[n]|,k) \in \mathbb{R}^{Q\times 1}$ reads

$$\boldsymbol{\xi}_n(|x[n]|,k) = \begin{cases} 1 & \text{if } k = p_n \\ 0 & \text{if } k \neq p_n \end{cases}, \text{ for } k = 1, 2, \cdots, Q, \tag{5}$$

where k indicates the index within the vector, and p_n is defined as

$$p_n = \left\lfloor \frac{|x[n]|}{\Delta_x} \right\rfloor + 1. \tag{6}$$

Thus, the input sample $x[n]$ is multiplied with the corresponding LUT gain, which is indexed by the input magnitude $|x[n]|$. Additionally, Δ_x is the amplitude spacing of the LUT entries, defined as the maximum input magnitude divided by the desired number of LUT entries, Q.

In this paper, we specifically utilize the so-called injection-based DPD scheme, in which an estimate of the PA nonlinear distortion products is injected, properly phased, to the linear digital signal such that the PA output signal is effectively linearized. Following this scheme, we rewrite the final form of the input-output relation of the DPD model, illustrated in Fig. 2, as

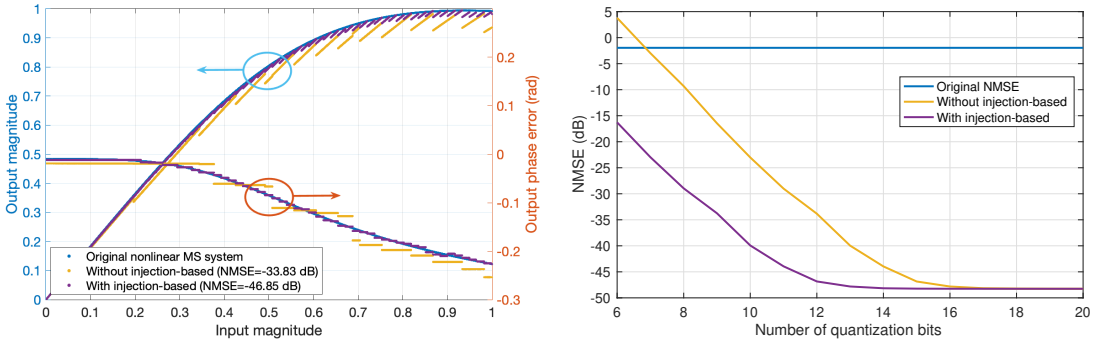
$$x_{\mathrm{DPD}}[n] = x[n] + \sum_{m=0}^{M-1} x[n-m]\boldsymbol{\xi}_{n-m}^T(|x[n-m]|,k)\mathbf{c}_m. \tag{7}$$

Applying such formulation will reduce the dynamic range of the LUT entries, thus requiring less number of bits in fixed-point implementations. This reduction is further explored and analyzed in the Subsection II.C.

The obtained input-output relation of the predistorter can be now equivalently expressed in matrix notation, for an N-sized block of samples, as

$$\mathbf{x}_{\mathrm{DPD}} = \mathbf{x} + \mathbf{\Phi}\mathbf{w}, \tag{8}$$

where $\mathbf{\Phi} \in \mathbb{C}^{N\times C}$ is the input data basis functions matrix, whose structure is shown in (1), with $C = MQ$ being the total number of model coefficients, and $\mathbf{x} = [x[n], x[n+1], \cdots, x[n+N-1]]^T$ denotes the input data vector. The column $\mathbf{w} \in \mathbb{C}^{C\times 1}$ stacks the M LUTs (i.e. $\mathbf{c}_0,\ \mathbf{c}_1, \cdots, \mathbf{c}_{M-1}$) to form the complete set of DPD coefficients, and it is typically initialized as a zero vector in the first DPD iteration.



(a) (b)

Fig. 3. Direct modeling performance of the modified Saleh (MS) model with and without the injection-based scheme, when considering a fixed number of quantization bits, shown in (a), and when varying the number of quantization bits, shown in (b).

B. Closed-Loop Learning – Unsigned Algorithms

Formulating the LUT-based DPD as a linear-in-parameters model as in (8), allows us to apply closed-loop learning techniques. Defining the error signal $\mathbf{e}_k \in \mathbb{C}^{N\times 1} = \mathbf{x}_k - \frac{\mathbf{y}_k}{G_{\text{PA}}}$, for block iteration k, we can define the three learning techniques which are studied along this paper, namely the damped Gauss-Newton (GN), the self-orthogonalized (SO), and the block-LMS (BLMS) [21], [23], [27] methods. These learning approaches can be expressed as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_g \left(\mathbf{\Phi}_k^H \mathbf{\Phi}_k\right)^{-1} \mathbf{\Phi}_k^H \mathbf{e}_k, \tag{9}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_s \mathbf{R}^{-1} \mathbf{\Phi}_k^H \mathbf{e}_k, \tag{10}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_b \mathbf{\Phi}_k^H \mathbf{e}_k, \tag{11}$$

where μ_g, μ_s and μ_b are the corresponding learning step-sizes for each method. Additionally, $\mathbf{R}$ is the covariance matrix of the input basis function vector, formally defined as

$$\mathbf{R} = \mathrm{E}[\mathbf{\Phi}_n \mathbf{\Phi}_n^H], \tag{12}$$

where

$$\mathbf{\Phi}_n = [x[n]\boldsymbol{\xi}_n^T(|x[n]|,k)\ x[n-1]\boldsymbol{\xi}_{n-1}^T(|x[n-1]|,k)\cdots \\ \cdots\ x[n-M+1]\boldsymbol{\xi}_{n-M+1}^T(|x[n-M+1]|,k)]^T. \tag{13}$$

The matrix $\mathbf{R}$ can be precomputed and fixed, and thus its online calculation is not required [31].

Finally, we note that the formulations in (8)–(12) are quite general, and can be applied with other LUT-based DPD models as well, such as those following generalized MP or Volterra-DDR models (see [23] for an example).

C. Dynamic Range Reduction in $\mathbf{q}$

In order to shortly assess and illustrate the dynamic range reduction in the LUT control points $\mathbf{q}$ through the injection-based processing principle, the modified Saleh (MS) model presented in [32] is considered as a practical and reproducible example. This MS model is approximated with and without the injection-based DPD scheme, in other words, the processing principles in (7) and (3) are deployed, but here in the context of direct PA modeling instead of DPD. The results are then compared to demonstrate the benefit of using the injection-based approach. For clarity, the AM-AM and AM-PM responses of the considered MS model are stated as [32]

$$z(r) = \frac{\alpha_z r}{\sqrt{1+\beta_z r^3}}, \tag{14}$$

$$\psi(r) = \frac{\alpha_\psi}{\sqrt[3]{1+\beta_\psi r^4}} - \epsilon, \tag{15}$$

where r and z represent the instantaneous input and output envelope values, while ψ represents the output signal phase change as a function of the input envelope. Furthermore, $\alpha_z = 0.82$, $\beta_z = 0.29$, $\alpha_\psi = -0.35$, $\beta_\psi = 1$, and $\epsilon = -0.36$ are the envelope and phase related model coefficients, which have been estimated in this case from the measured input-output relation of an LDMOS PA, used in [32].

A 100 MHz 5G NR compliant OFDM signal, of length 20.000 samples, with 30 kHz subcarrier spacing (SCS), 273 active resource blocks (RB), and 7 dB PAPR measured at 0.01% point of the complementary cumulative distribution function (CCDF) is then generated and passed through the model. Next, an LS fitting technique is used to estimate the LUT control points modelling the MS PA, with and without the injection-based scheme. Both LUT vectors are then quantized to the same number of bits (12 quantization bits), and the modelling capabilities of both approaches are then visually illustrated in Fig. 3a. Additionally, the direct modeling related NMSE numbers are calculated and presented as the number of quantization bits increases from 6 to 20. These NMSE results are shown in Fig. 3b. It can be observed that the modelling accuracy or performance is steadily about 13 dB better when considering the injection-based scheme, until the NMSE values essentially saturate with 16 quantization bits. This illustrates and quantifies the dynamic range reduction obtained through the injection-based approach.

D. Observation Receiver and Feedback in mmW Active Array Systems

There are generally several alternatives to address the observation receiver (ORX) aspect and arranging the feedback signal for DPD parameter learning in mmW active array systems. One known method is the hardware-based approach where the individual PA output signals are phase-aligned and combined in hardware [7], [10] – illustrated conceptually also in Fig. 1. Another alternative is to adopt a separate ORX to capture the over-the-air (OTA) combined signal [9], [13], [33]–[35] and feed it back to the transmitter system through some means for DPD learning. Also this alternative approach is illustrated in Fig. 1. Both of these approaches basically seek to mimic the far-field signal at the actual receiver, under the assumption of line-of-sight propagation.

We clarify that the DPD learning algorithms proposed in this article do not explicitly depend on the actual method of obtaining the combined observed signal, while note that the hardware-based ORX system has the benefit of, e.g., avoiding the OTA ORX positioning and beam misalignment challenges [13], [36]. For fairness, however, it is also noted that there exists literature, e.g., [33]–[35], where the OTA ORX beam misalignment challenge is further addressed, proposing different mechanisms to reconstruct the far-field signal in the direction of the main beam, through subsequent sidelobe observations or by leveraging the crosstalk between adjacent radiating elements. In our actual mmW active array experiments in Section V, our feedback system adopts a carefully aligned OTA ORX with the primary purpose of mimicking the hardware-based feedback combiner system.

III. Signed Learning Algorithms

In order to reduce the computational complexity of the baseline learning rules in (9), (10), and (11), we next formulate computationally efficient sign-based learning algorithms. In general, the idea behind the signed algorithms is to sign selected terms in the learning equations, such that the needed number of multiplications is largely reduced. This is beneficial as in the digital signal processing (DSP) implementations, multiplications constitute one of the most resource-intensive operations, while additions are essentially free [4], [37].

The classical definition of the complex signum function projects a non-zero complex number to the unit circle in the complex plane [38]. The magnitude of the resulting number, $\bar{z}$, is 1, but the real and imaginary parts are not equal to ± 1, thus no direct complexity reduction can yet be achieved when multiplying with $\bar{z}$. To remove the need for multiplications, we define the complex signum function instead as

$$\mathrm{csgn}(z) := \mathrm{sgn}(\mathrm{Re}(z)) + j\,\mathrm{sgn}(\mathrm{Im}(z)), \tag{16}$$

which provides either -1 or $+1$ for the real and imaginary parts. For matrices, the operation is taken element-wise. The next sections present the three considered signed algorithms, implemented with the form shown in (16), and its combination with the original learning equations.

A. The Sign Algorithm

The sign algorithm is obtained by signing the error signal $\mathbf{e}_k$ in the learning rules presented in (9), (10), and (11). With this simplification, multiplications in the term $\mathbf{\Phi}_k^H \mathbf{e}_k$ are avoided. We note that the dimension of $\mathbf{e}_k$ is N – commonly a large number in DPD implementations (in the experiments of this paper, $N = 25,000$) – thus a large reduction in terms of multiplications can be achieved. By signing the error vector, the DPD learning rules read

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_g \big(\mathbf{\Phi}_k^H \mathbf{\Phi}_k\big)^{-1} \mathbf{\Phi}_k^H\, \mathrm{csgn}(\mathbf{e}_k), \tag{17}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_s \mathbf{R}^{-1} \mathbf{\Phi}_k^H\, \mathrm{csgn}(\mathbf{e}_k), \tag{18}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_b \mathbf{\Phi}_k^H\, \mathrm{csgn}(\mathbf{e}_k). \tag{19}$$

The reader can find an implementation of the sign algorithm in combination with GN learning rule in [20].

B. The Signed Regressor Algorithm

The SRA method signs the transposed basis functions matrix, $\mathbf{\Phi}_k^H$, in the learning rules. Hence, multiplications in the terms $\mathbf{\Phi}_k^H \mathbf{\Phi}_k$ and $\mathbf{\Phi}_k^H \mathbf{e}_k$ (GN), and $\mathbf{\Phi}_k^H \mathbf{e}_k$ (BLMS) are avoided, making the computational complexity of the learning rule lighter. In the GN method, the complexity saving is larger compared to the Sign algorithm, as an extra term is signed in the learning equation. In the BLMS method, no reduction is achieved when compared to the previous Sign algorithm. The SRA learning rules corresponding to GN and BLMS methods can be expressed as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_g \Big(\mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{\Phi}_k\Big)^{-1} \mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{e}_k, \tag{20}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_b\, \mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{e}_k. \tag{21}$$

When referring to the SO method, the SRA approach cannot be applied as such, as the inverse covariance matrix, $\mathbf{R}^{-1}$, already contains the input data matrix multiplication. Two alternative solutions can be drawn in order to use the SRA principle in combination with SO learning rule. The first proposed form signs only the input data matrix term $\mathbf{\Phi}_k^H$, avoiding the calculation of $\mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{e}_k$. The second form signs the inverse covariance matrix $\mathbf{R}^{-1}$, simplifying the matrix multiplication $\mathrm{csgn}(\mathbf{R}^{-1})\mathbf{\Phi}_k^H$. Thus, the exact learning rules can be expressed as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_s \mathbf{R}^{-1}\, \mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{e}_k, \tag{22}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_s\, \mathrm{csgn}(\mathbf{R}^{-1})\mathbf{\Phi}_k^H \mathbf{e}_k. \tag{23}$$

It is noted that with the former formulation, the computational complexity is the same as in the Sign SO case, since an equal number of multiplications is avoided.

It is also important to note that all polynomial-based DPD approaches, as well as linearly interpolated LUTs, basically suffer from a rank deficiency in the signed data matrix $\mathrm{csgn}(\mathbf{\Phi}_k^H)$, as repeated columns or linear combinations between them will appear. An example is presented in [21], in the context of an MP DPD [39]. In such a case, the estimated DPD coefficients will diverge, as they do not have a unique solution. One way to solve this problem is to apply a unitary Walsh-Hadamard transformation (WHT) to gaussianize the distribution of $\mathrm{csgn}(\mathbf{\Phi}_k^H)$ and make it full rank [21]. This, however,

TABLE I
COMPLEXITY ANALYSIS OF THE BASELINE AND THE SIGNED LEARNING METHODS FORMULATED AND ADOPTED IN THE PAPER, AS A FUNCTION OF THE MODEL PARAMETERS, IN TERMS OF REAL MULTIPLICATIONS AND REAL ADDITIONS PER DPD LEARNING ITERATION.

	Real multiplications	**Real additions**
Gauss-Newton	$C^3 + 4M^2(N+1) + 2M(2N+1)$	$2M^2(N-1) + 2M(N+M-2) + 2C$
Sign Gauss-Newton	$C^3 + 4M^2(N+1) + 2C$	$2M^2(N-1) + 2M(N+M-2) + 2C$
SRA Gauss-Newton	$C^3 + 4M^2 + 2M$	$2M^2(N-1) + 2M(N+M-2) + 2C$
Sign-sign Gauss-Newton	$C^3 + 2M$	$2M^2(N-1) + 2M(N+M-2) + 2C$
Self-orthogonalization	$4(MN + M^2) + 2C$	$2M(N+C-2) + 2C$
Sign self-orthogonalization	$4CM + 2C$	$2M(N+C-2) + 2C$
SRA 1 self-orthogonalization	$4CM + 2C$	$2M(N+C-2) + 2C$
SRA 2 self-orthogonalization	$4MN + 2C$	$2M(N+C-2) + 2C$
Sign-sign 1 self-orthogonalization	$2C$	$2M(NM+1)$
Sign-sign 2 self-orthogonalization	$2C$	$2M(N+C-2) + 2C$
Block-LMS	$2M(2N+1)$	$2(MN+C)$
Sign block-LMS	$2M$	$2(MN+C)$
SRA block-LMS	$2M$	$2(MN+C)$
Sign-sign block-LMS	0	$2(MN+C)$

further increases the complexity in the learning rule. On the other hand, and very importantly, with the proposed LUT-based DPD approach, the rank deficiencies are avoided, as the structure of this model does not lead to repeated or linearly dependent columns in $\mathrm{csgn}(\mathbf{\Phi}_k^H)$. Thus, the SRA learning rule can be directly applied, with no extra matrix transformations needed. This is one clear benefit of the proposed LUT-based DPD formulation compared to polynomial based DPDs.

C. The Sign-Sign Algorithm

Finally, the Sign-sign algorithm applies the signum function to both the data matrix and the error vector. In the GN method, the required multiplications are greatly reduced, as only a few matrix operations need to be calculated. In the case of the BLMS approach, the number of required multiplications to obtain the DPD coefficients is already zero. The exact learning expressions with the Sign-sign algorithm for GN and BLMS approaches read

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_g \Big(\mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{\Phi}_k\Big)^{-1} \mathrm{csgn}(\mathbf{\Phi}_k^H)\,\mathrm{csgn}(\mathbf{e}_k), \tag{24}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_b\,\mathrm{csgn}(\mathbf{\Phi}_k^H)\,\mathrm{csgn}(\mathbf{e}_k). \tag{25}$$

When referring to the SO method, the same conclusion as presented with the SRA algorithm is drawn, i.e., the Sign-sign approach cannot be directly applied to the learning rule. We thus define again two revised alternative solutions for the Sign-sign algorithm in combination with the SO, which read

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_s \mathbf{R}^{-1}\,\mathrm{csgn}(\mathbf{\Phi}_k^H)\,\mathrm{csgn}(\mathbf{e}_k), \tag{26}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu_s\,\mathrm{csgn}(\mathbf{R}^{-1})\mathbf{\Phi}_k^H\,\mathrm{csgn}(\mathbf{e}_k). \tag{27}$$

With both solutions, the DPD learning complexity, in terms of real multiplications, is almost reduced to zero, while the exact computational complexity assessment is provided in the next section.

Additionally, it is noted that the same discussion about the rank deficiency problem in $\mathrm{csgn}(\mathbf{\Phi}_k^H)$ propagates with the Sign-sign algorithm as well, with respect the polynomial-based and interpolated LUT DPD approaches. In other words, the LUT-based DPD formulation can be used without any additional matrix transformations as there are no rank-deficiency challenges.

IV. LEARNING COMPLEXITY ANALYSIS AND COMPARISON

The DPD learning complexity is analyzed in terms of real multiplications and real additions per DPD coefficient update, over an N-sized block of samples. It is assumed that one complex multiplication is implemented with 4 real multiplications and 2 real additions, and one real-complex multiplication costs 2 real multiplications. Furthermore, one complex addition costs 2 real additions, while a real-complex addition is performed with one addition. In the complexity assessment, when it comes to matrix algebra, we follow [40].

Firstly, Table I presents the complexity expressions of the GN, SO, and BLMS adaptive learning methods, as functions of the DPD model parameters. These expressions essentially cover the original learning rules presented in (9), (10), and, (11) and the sign-based versions presented in Section III. Secondly, Table II shows example numerical complexity numbers, with $N = 25,000$ samples, $Q = 32$, $M = 4$, and $C = MQ = 128$, which represent the same parametrization used in the experimental measurement results in Section V. Additionally, this table illustrates the complexity percentage reduction of the sign algorithms with respect to the original learning equations, in terms of real multiplications. As seen herein, the number of real multiplications is commonly very largely reduced when deploying the sign-based, thus greatly easing continuous learning and/or on-chip learning implementations.

Several concluding remarks can be extracted from the complexity analysis, as follows:

TABLE II
NUMERICAL COMPLEXITY ASSESSMENT AND COMPARISON OF THE METHODS WHEN $N = 25,000$, $Q = 32$, $M = 4$, AND $C = MQ = 128$. THE RELATIVE COMPLEXITY REDUCTION WITH RESPECT TO THE ORIGINAL LEARNING EQUATIONS IS ALSO SHOWN, IN PERCENTAGES.

	Real multiplications	**Mul. reduction**	**Real adds.**
GN	4×10^6	0%	1×10^6
Sign GN	3.7×10^6	7.5%	1×10^6
SRA GN	2×10^6	50%	1×10^6
Sign-sign GN	2×10^6	50%	1×10^6
SO	401×10^3	0%	201×10^3
Sign SO	2.3×10^3	$> 99\%$	201×10^3
SRA 1 SO	2.3×10^3	$> 99\%$	201×10^3
SRA 2 SO	400×10^3	$< 1\%$	201×10^3
Sign-sign 1 SO	256	$> 99\%$	800×10^3
Sign-sign 2 SO	256	$> 99\%$	201×10^3
BLMS	400×10^3	0%	200×10^3
Sign BLMS	8	$> 99\%$	200×10^3
SRA BLMS	8	$> 99\%$	200×10^3
Sign-sign BLMS	0	100%	200×10^3

- GN SRA and Sign-sign algorithms pose mutually the same complexity order of magnitude, thus the model which provides better performance should be selected.
- SO Sign and SRA 1 algorithms pose mutually the same complexity, thus the model which yields better performance should be selected. The same conclusion applies for Sign-sign 1 and Sign-sign 2 SO.
- SO SRA 2 model does not provide essentially any complexity reduction, hence it is not explicitly considered in the measurement based experiments.
- Sign and SRA BLMS algorithms pose mutually the same complexity, thus the model which provides better performance should be selected.

All in all, the sign algorithms are capable of drastically reducing the computational complexity, especially in the cases of BLSM and SO, where in most cases it is reduced by more than 99% with respect to the corresponding original update rules. In the case of GN, the sign algorithms simplify the update up to 50%. In the next section, the DPD linearization performance of the proposed algorithms will be evaluated through extensive RF measurements. Together with the complexity analysis, it will allow to assess the complexity-performance trade-offs of the proposed algorithms.

V. EXPERIMENTAL RESULTS

In order to test and validate the proposed DPD algorithms, extensive set of experimental results is provided building on mmW FR-2 OTA measurements. Specifically, our setup features a state-of-the-art 28 GHz active antenna array with 64 integrated PAs and antenna units, with which the linearization performance-complexity trade-offs of the injection-based closed-loop MP-LUT DPD system are assessed and pursued, while deploying and comparing both the baseline and the various signed learning rules.

In the context of mmW array measurements, some important issues are to be noted. Firstly, an active antenna array with K antenna units contains also K parallel PA units. Furthermore, the different PA units are commonly mutually different, at least to certain extent, thus each parallel PA has unique nonlinear characteristics. Hence, the estimated predistorter building on combined observation path and combined observed signal can typically provide good linearization mostly in the array's main beam direction, while the beampattern of the array will maintain the levels of OOB distortion sufficiently low in other directions [7]. Secondly, the load modulation of the PAs, which occurs due to the coupling between the antennas [9], makes the effective nonlinear characteristics of the array *beam-dependent*. This essentially means that the optimal DPD solution will depend on the beam direction, and thus, the linearization solutions should take this into account. Real-time tracking and fast adaptive DPD learning are viable solutions, capable of estimating and adapting the DPD coefficients as the beam is steered. Third, the frequency selectivity of the transmitter system and thus that of the nonlinear distortion can already be substantial – mostly due to the wide channel bandwidths at mmW frequencies, calibration challenges and more difficult impedance matching, compared to lower frequencies.

In this work, the in-band DPD linearization performance is evaluated through the well-known error vector magnitude (EVM) metric [4], [14]. Additionally, since an OTA DPD system is considered, the out-of-band performance is measured with the total radiated power (TRP) based adjacent channel leakage ratio (ACLR), which is the ratio of the filtered mean power centered on the assigned channel frequency and the filtered mean power centred on an adjacent channel frequency, measured by integrating the powers over the whole beamspace, while keeping the *beamforming angle* fixed [14].

A. 28 GHz Active Array Experimental Setup

The OTA FR-2 measurement setup is depicted in Fig. 4. The transmit chain consists firstly of a Keysight M8190 arbitrary waveform generator (AWG), outputting the I/Q samples at 3.5 GHz IF. Then, a Keysight N5183B-MXG signal generator, providing the LO signal at 24.5 GHz, and a Marki Microwave T31040 mixer, further upconvert the signal to 28 GHz, after which the signal is filtered by a Marki Microwave FB3300 band-pass filter (BPF) to suppress the mixer induced image frequencies. Two preamplifiers, Analog Devices HMC499LC4 and Analog Devices HMC1131, are then deployed before the actual active antenna array to facilitate driving the array towards saturation. The test device is a 64-element Anokiwave AWMF-0129 antenna array, which transmits and radiates the signal OTA. It is mounted on an electrical tripod capable of providing the horizontal rotation, with 0 degrees considered as the array *beamforming angle* in these measurements. The radiated signal is then captured by a horn antenna and an observation receiver, such that the receiver antenna is well-aligned with the transmitter main beam. The observed signal

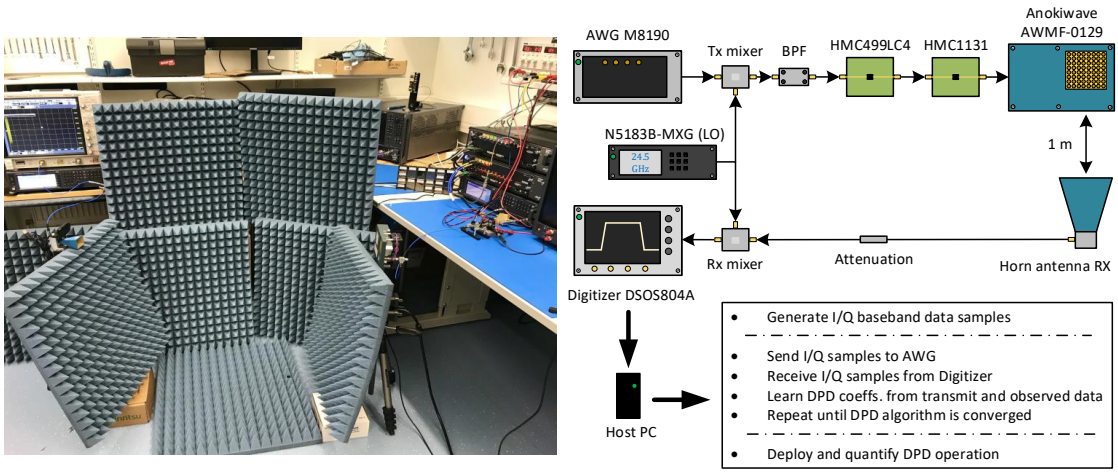


Fig. 4. 5G NR FR-2 OTA RF measurement setup utilized in the mmW DPD experiments.

is attenuated and downconverted again to IF frequency by another mixing stage. Finally, the resulting signal is fed into a Keysight DSOS804A oscilloscope, which is used as the actual digitizer to facilitate the post-processing on a host PC, where the DPD algorithms are executed. It is noted that the OTA RX with horn antenna is used both for DPD learning, as ORX, and for final OTA measurements to assess the DPD performance. The use of the OTA RX as the ORX is because the AWMF-0129 active array does not allow for actual hardware-based combiners for feedback, hence we deliberately mimic such through the carefully aligned OTA ORX.

The signals adopted in the coming Sub-sections B-G are 3GPP 5G NR Release-15 FR-2 compliant OFDM waveforms, with 120 kHz SCS and 264 RBs. This configuration maps to the channel bandwidth configuration of 400 MHz [14]. The signals adopted in Sections H-I are, in turn, generated by doubling the number of active subcarriers and the OFDM waveform processing FFT size, compared to the standard-compliant signal, which then maps already to an impressive channel bandwidth of 800 MHz. This is done deliberately to experiment and demonstrate the DPD-based active array linearization with extremely large channel bandwidths and OFDM modulation, while operating with effective isotropic radiated powers (EIRPs) of more than +40 dBm – something that has not been commonly reported in the existing literature.

In all experiments, the initial PAPR of the digital waveform is 9.5 dB, when measured at the 0.01% point of the instantaneous PAPR CCDF, and is then limited to 7 dB through well-known iterative clipping and filtering based processing, while also additional time-domain windowing is applied to suppress the inherent OFDM signal sidelobes. These impose an EVM floor of some 4% to the transmit signal. In a single DPD iteration, a block of $N = 25,000$ pseudo-random samples of the above-described 5G NR OFDM waveforms is circularly transmitted, received, and used to update the DPD coefficients. A new block of N samples is then generated for the next DPD update iteration. This transmission/reception and the DPD update is repeated until the DPD learning algorithm reaches convergence. The MP-LUT DPD models utilize LUT entry sizes of $Q = 32$, and $M = 4$ memory branches, as the baseline. The LUTs are initialized as all-zero vectors in the first DPD iteration. In the measurements where the SO learning rule is considered, the covariance matrix, $\mathbf{R}$, is estimated from a long sequence of 10 Msamples, and inverted before the actual DPD processing. It is then kept fixed during the remaining DPD iterations. A classical MP model in a closed-loop configuration, with $P = 11$ and $M = 4$, is utilized as the reference method, as polynomial based DPDs are some of the most common high-performance techniques used in the literature [4], [16], [41]. Furthermore, the parameter learning of the MP reference method builds always on the unsigned GN algorithm as the polynomial basis functions are known to be largely correlated and here no basis function orthogonalization is adopted.

B. DPD Performance

In this subsection, the OTA DPD linearization performance of the various GN, SO, BLMS, and their corresponding sign versions, is demonstrated. The following measurements are carried out with the NR FR-2 400 MHz signal, measured at a highly nonlinear operation point of the active antenna array, specifically at EIRP of approximately +43 dBm.

Firstly, the measured power spectral densities (PSDs) corresponding to the GN learning rule are presented in Fig. 5. The performance of the sign algorithms is observed to be very close to that of the classical unsigned learning method, despite the substantially reduced complexity. Even the highly simple Sign-sign algorithm provides a comparable linearization performance to the original learning rule. It is also observed that

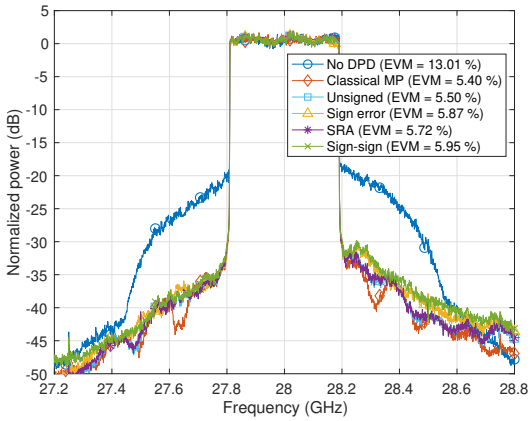


Fig. 5. 400 MHz 5G-NR OTA linearization performance of the closed-loop MP-LUT DPD, at EIRP of +43 dBm, with original GN and signed GN learning algorithms. Also the performance of classical MP DPD with unsigned learning is shown, for reference.

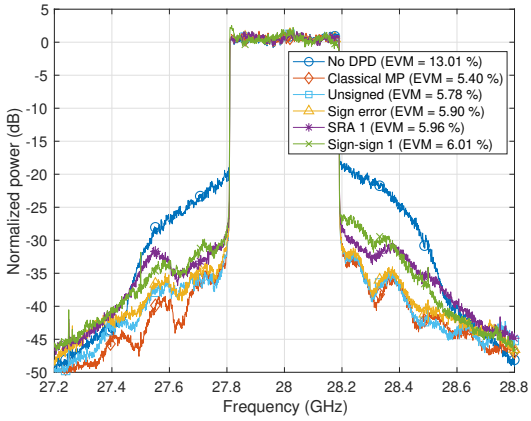


Fig. 6. 400 MHz 5G-NR OTA linearization performance of the closed-loop MP-LUT DPD, at EIRP of +43 dBm, with original SO and signed SO learning algorithms. Also the performance of classical MP DPD with unsigned GN learning is shown, for reference.

the linearization results are very close to the reference MP model.

Secondly, the measured PSDs corresponding to the SO learning are presented in Fig. 6. In this case, somewhat decreased linearization performance is expected, and also observed, compared to GN as the learning equation applies a fixed estimated covariance matrix $\mathbf{R}$. However, it also involves further reduced complexity. It can be observed that the unsigned and Sign SO achieve mutually similar linearization performance, also been very close to the classical MP. These are then followed in performance by the SRA 1 and Sign-sign 1 algorithms. As seen in Section IV, the sign error and SRA 1 achieve the same complexity reduction, so the model providing

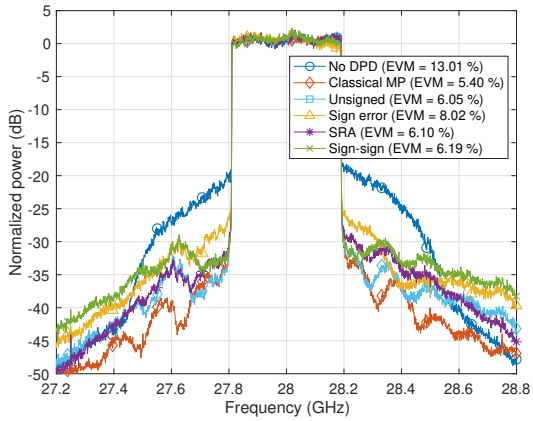
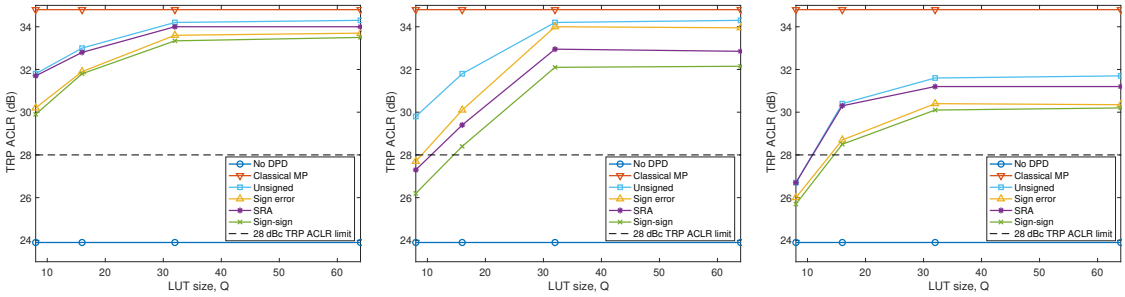


Fig. 7. 400 MHz 5G-NR OTA linearization performance of the closed-loop MP-LUT DPD, at EIRP of +43 dBm, with original BLMS and signed BLMS learning algorithms. Also the performance of classical MP DPD with unsigned GN learning is shown, for reference.

better performance should be selected – in this case, the Sign error approach. It is also important to note that the complexity of the Sign-sign algorithm is close to zero, while still achieving a fair amount of linearization. When measuring with the Sign-sign 2 method, the DPD does not converge to any reasonable solution, thus the Sign-sign 1 method is deployed from now on when it comes to the Sign-sign based approaches.

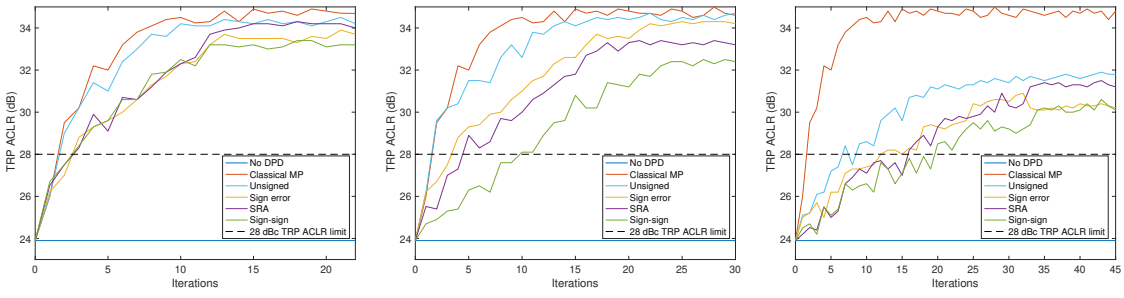
Thirdly, Fig. 7 presents the measured PSDs with the BLMS method. In this case, the unsigned and the SRA learning approaches provide the best performance. These provide again performance fairly similar to the classical MP case, while being a bit more degraded when compared to GN and SO cases. The Sign and Sign-sign curves follow somewhat behind, in performance, but also facilitate good linearization despite no actual multiplications are needed in the parameter learning.

Overall, it can be observed that the best performance is achieved with the GN learning methods, however, GN learning also involves the highest computational complexity. The SO approach presents an intriguing solution, able to provide very similar levels of linearization as GN, with clearly reduced complexity. The BLMS learning approach is, in turn, the simplest method in terms of complexity, and also capable of facilitating good amounts of linearization. Additionally, all methods essentially reach the 8% EVM requirement [14] of NR Release-15 that corresponds to 64-QAM – the largest modulation order supported currently at FR-2. From the complexity-performance trade-off point of view, we observe that the SO Sign-sign 1, BLMS Sign, BLMS SRA, and BLMS Sign-sign are particularly interesting as they require exactly or approximately zero multiplications per DPD iteration, while still providing good linearization performance. Also, it is noted and emphasized that the BLMS and different signed BLMS variants are indeed applicable with the injection-based MP-LUT DPD – without any additional orthogonalization proce-



(a) (b) (c)

Fig. 8. Comparison of the LUT size, Q, in the closed-loop MP-LUT DPD vs. measured TRP ACLR for (a) GN, (b) SO, and (c) BLMS DPD learning rules at EIRP of +43 dBm. The 28 dB ACLR limit is also shown, together with MP DPD reference performance with unsigned GN learning. These results are obtained by training the DPD model with the same number of iterations as presented in the x-axes of Fig. 9.



(a) (b) (c)

Fig. 9. TRP ACLR convergence of the closed-loop MP-LUT DPD vs. number of block-iterations with (a) GN, (b) SO, and (c) BLMS algorithms at EIRP of +43 dBm. The 28 dB ACLR limit is also shown, together with MP DPD reference convergence with unsigned GN learning.

dures that are commonly adopted in case of, e.g., gradient-adaptive canonical MP DPD [27]. This is a clear benefit compared to polynomial based DPD systems.

C. LUT Entry Size Comparison

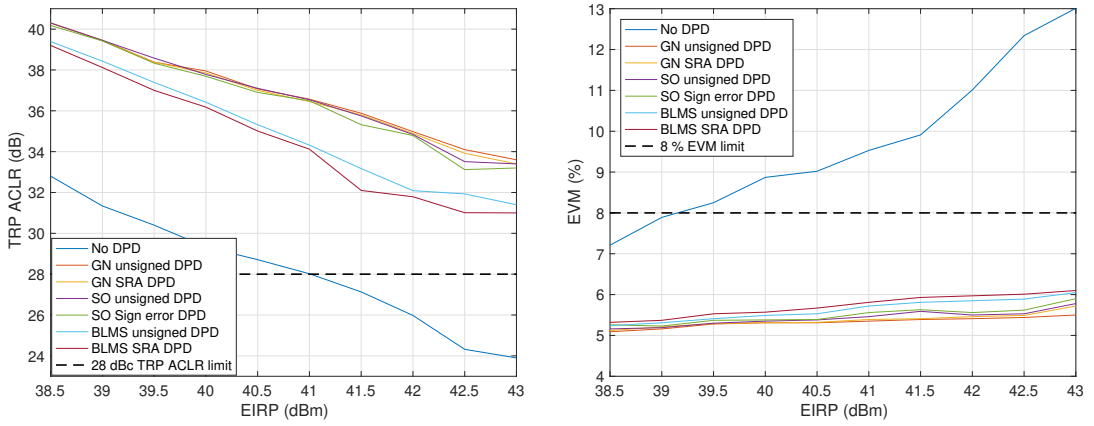
We next continue the OTA measurements with NR FR-2 400 MHz signal at EIRP of +43 dBm while now varying the LUT entry size, Q, in the proposed closed-loop MP-LUT DPD method to experiment and assess its impact on the linearization performance. The obtained measured results with GN, SO, and BLMS are presented in Fig. 8. As can be observed, when adopting the DPD models with small numbers of entries in the LUT, the performance drops to some extent. This is quite expected as very few control points in the LUT are not sufficient to accurately model and invert the effective PA nonlinearity. At the same time, it is observed that as the LUT entry size is increased, the TRP ACLR performance improves, until reaching $Q = 32$, at which the performance essentially saturates in these measurements. Compared to the results in [23], the considered injection-based DPD scheme allows for lower entry-sized non-interpolated LUTs, while the sign methods further reduce the DPD processing and learning complexities. Additionally, we observe that the 5G NR TRP ACLR limit of 28 dBc [14] is fulfilled in all cases when $Q = 16$ or greater.

This experiment also reconfirms the conclusion drawn in the previous subsection, showing that the linearization performances of the original and selected signed algorithms are very close to each other. Specifically, the difference is only 0.1 dB between GN unsigned and GN SRA, 0.2 dB between SO unsigned and SO Sign, and 0.4 dB between BLMS unsigned and BLMS SRA. The computational complexity, in turn, is reduced by 50% in the first case, and by more than 99% with the SO and the BLMS, as analyzed and shown in Section IV.

D. DPD Convergence

We next pursue and present the convergence behavior of the proposed DPD solutions, with the same configuration as adopted before. The convergence behavior is presented in terms of the measured OTA TRP ACLR as a function of the number of DPD block-iterations, again with a block-length of $N = 25,000$ samples.

The obtained convergence results are presented in Fig. 9, for the GN, SO, and BLMS learning rules. In general, it is observed that the convergence speed is faster with the unsigned versions of the learning equations, reaching the steady-state sooner. When applying the signed algorithms, the convergence



(a) (b)

Fig. 10. Measured TRP-based ACLR in (a), and EVM in (b), as functions of the EIRP in the 400 MHz channel bandwidth case at 28 GHz.

speed is then slightly decreased, as some information is lost when signing different terms in the learning rules. However, the signed solutions are capable of reaching the steady-state only in a few more iterations, as shown in Fig. 9. In all cases, the models are very stable after convergence, and fulfill the 28 dBc TRP ACLR limit [14].

The different algorithms are capable of reaching full convergence in around 10-12 (GN), 17-20 (SO), and 30-35 (BLSM) iterations, respectively. The relative behavior is intuitive, as the GN algorithm calculates the inverse covariance matrix, $(\mathbf{\Phi}_k\mathbf{\Phi}_k^H)^{-1}$, in each DPD iteration, thus providing the fastest convergence. The SO learning equation considers a fixed covariance matrix estimate, that somewhat slows down the convergence. The BLMS, in turn, can be interpreted to consider an identity covariance matrix, which is already a very crude approximation, thus the convergence is slowest and also the steady-state performance is somewhat lower.

E. Power Sweep

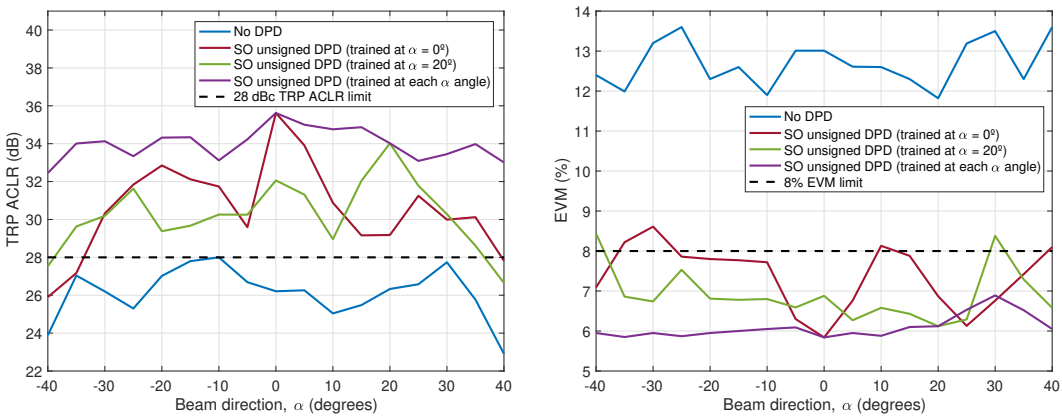
This fourth experiment considers a transmit power sweep carried out with the same configuration as presented above, and illustrates the measured TRP ACLR and EVM values as functions of the EIRP. By sweeping the EIRP, two main things can be studied. First, to evaluate whether the EVM or the TRP ACLR is the limiting performance metric of the system [14], in terms of the maximum EIRP. Second, to assess the performance of the DPD algorithms as the array output power varies. In this study, GN (unsigned and SRA), SO (unsigned and Sign), and BLMS (unsigned and SRA) algorithms are chosen and measured, as they have been observed in the earlier examples to have particularly positive performance-complexity trade-offs.

The measured TRP ACLR and EVM values as functions of the EIRP are presented in Fig. 10. Firstly, it can be clearly seen that, when no DPD is applied, the EVM constitutes the metric limiting the maximum achievable EIRP, such that both TRP ACLR and EVM are still fulfilled. Specifically, when no DPD is applied, the EIRP is limited to some +39.2 dBm, while when DPD processing is utilized, both requirements are still fulfilled at least up until +43 dBm, and clearly also somewhat beyond. These findings indicate a power efficiency increase in the overall transmitter, as the antenna array can be operated closer to saturation thanks to the transmit power increase facilitated by the DPD operation.

Secondly, it can be seen that the DPD algorithms behave in a similar manner as concluded in earlier subsections. The best linearization performance is obtained with GN and its signed version. The linearization performance obtained with unsigned and sign SO lies very close to GN, despite the reduced learning complexity. The BLMS follows somewhat behind, providing less linearization performance, but constituting a very simple DPD solution. In general, the sign algorithms lie very close to the original learning rules, and allow for complexity reductions up to 50% (GN), and more than 99% (SO and BLMS). Additionally, as already noted, all the algorithms successfully fulfill the 3GPP specifications [14] at least up to EIRP of +43 dBm.

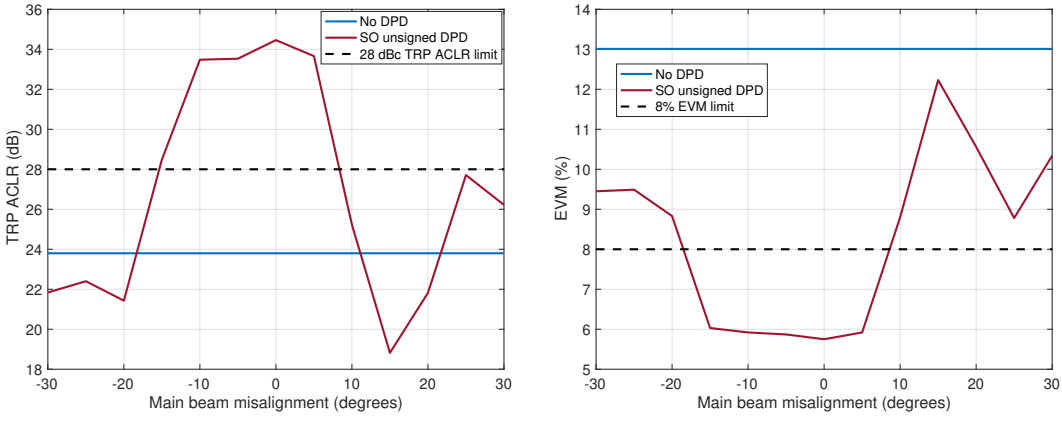
F. Beam-Dependence of Radiated Nonlinear Distortion

We next explore the effects of beam-steering on the nonlinear characteristics of the active array, while continue to utilize the same 5G NR OFDM waveform as in the previous experiments. Furthermore, for presentation simplicity, we focus only on the SO unsigned DPD learning method in this experiment. In these measurements, the transmit and receive antenna systems were kept at the same physical positions throughout the experiment, and were first aligned at $\alpha = 0°$ to estimate the DPD coefficients with the beam of the antenna array pointing towards this direction. Then, the electrical beam of the active array was digitally steered, sweeping from $\alpha = -40°$ to $\alpha = 40°$ with an angular resolution of $5°$, by means of phase-only analog beamforming. At the same time,



(a) (b)

Fig. 11. Linearization performance in terms of (a) TRP ACLR, and (b) EVM as a function of the steering angle at EIRP of +43 dBm, when the DPD is trained such that the main beam is aligned towards $\alpha = 0°$ (red) or $\alpha = 20°$ (green). Additionally, the purple curve presents the performance when the DPD system is separately re-trained for each beam direction angle α.



(a) (b)

Fig. 12. Linearization performance as a function of the OTA ORX misalignment angle at EIRP of +43 dBm, in terms of (a) TRP ACLR, and (b) EVM. The DPD coefficients are learned when the beams are misaligned, while the DPD verification is done when the beams are aligned. In all cases, the electrical beam angle of the active array is kept fixed at 0°.

the active array was physically rotated towards the opposite direction, so that the main beam of the array was always pointing to the receiver horn antenna. The same experiment was also repeated by training the DPD coefficients when the beam was pointed towards $\alpha = 20°$ direction. Additionally, a third set of measurements was also conducted, such that the DPD coefficients were always learned for every considered electrical beam direction. In all cases, the EVM and the TRP based ACLR, defined in Section V, were evaluated.

The results of these experiments are presented in Fig. 11. Through the red and green DPD curves, it can be seen that excellent linearization performance is obtained, in terms of both TRP ACLR and EVM, at those specific directions at which the DPD model was trained (i.e., $\alpha = 0°$ and $\alpha = 20°$). Additionally, the DPD performance then fastly deteriorates when the beam is steered to another angle. This is caused by the load modulation phenomenon, modifying the effective non-linear characteristics of the array when the beam is steered [9], [13]. We can observe that for some specific beam directions, the TRP ACLR and EVM values are no longer satisfying the specified 3GPP limits [14]. Finally, the purple curves show that the linearization performance remains stable when the DPD is trained at every beam direction, providing relatively flat TRP ACLR and EVM curves for different beamforming angles. Overall, these results indicate that continuous DPD tracking may indeed be needed in systems where fast beam-steering or beam-switching is adopted.

G. OTA ORX Misalignment Study

While the main purpose of the OTA ORX is, in this article, to mimic the hardware combined based feedback system, we next shortly address the DPD performance as a function of the misalignment between the beam directions of the active array and an OTA ORX. The same 400 MHz 5G NR waveform as in the previous experiments is utilized, while for simplicity, we focus on the SO unsigned DPD learning method.

First, the TX and OTA ORX beams were accurately aligned, with the electrical beam configured to $0°$. Then, the active array was mechanically rotated from $-30°$ to $30°$ with an angular resolution of $5°$ to produce controlled misalignment, and the learned DPD coefficients were stored for each misalignment angle. Afterwards, the transmit and receive beams were again accurately aligned, and the DPD performance was assessed such that the main beam was pointing accurately towards the receive antenna. The EVM and TRP-ACLR were calculated and recorded for each misalignment angle.

The obtained results are presented in Fig. 12. As it can be observed, the DPD system achieves an excellent linearization performance when the misalignment angle ranges from some $-10°$ to $+7°$. However, as the misalignment angle increases beyond this range, the DPD system experiences a systematic loss of linearization performance due to the fact that the OTA ORX is then already observing clearly off the main beam. It is also noted that the considered active array has a minimum in its radiation patter at around $+15°$, which can be seen from the TRP ACLR and EVM values at this angle. In our view, this highlights the benefits of a hardware-based observation system, shown in Fig. 1, which is by design immune to any misalignment. Alternatively, the methods from [33]–[35] can be considered such that observing with the OTA ORX outside the transmitter main beam becomes more feasible.

H. DPD Evaluation with Aggregated 800 MHz Channel BW

The last two experiments study and assess the linearization performance of the proposed algorithms in an aggregated 800 MHz channel bandwidth case with 5G NR like OFDM signal. Compared to the previous examples, we utilize here double the amount of passband active subcarriers and FFT processing size, while keeping the SCS of 120 kHz. This first study presents the OTA linearization performance of the GN (unsigned and SRA), SO (unsigned and Sign), and BLMS (unsigned and SRA) algorithms, while measuring at EIRP of approximately +41.5 dBm.

A corresponding snapshot linearization performance example, in terms of measured spectra, is presented in Fig. 13. It can be observed that the GN and SO models, and their signed versions, perform similarly, achieving TRP ACLR numbers of around 35 dB, while the corresponding EVM numbers are shown in the figure. The BLMS algorithm follows somewhat behind, still achieving good amount of linearization, with TRP ACLR numbers being around 32 dB. It is also noted that the transmitter system poses already quite substantial memory effects in the 800 MHz bandwidth case. However, based on this example and the following EIRP sweep measurements, the injection-based MP-LUT can well handle such memory effects.

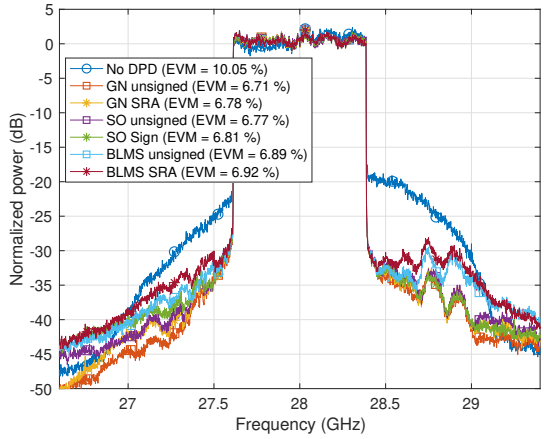


Fig. 13. 800 MHz 5G-NR like OTA linearization performance of the closed-loop MP-LUT DPD, at EIRP of +41.5 dBm, with unsigned and signed GN, SO, and BLMS algorithms.

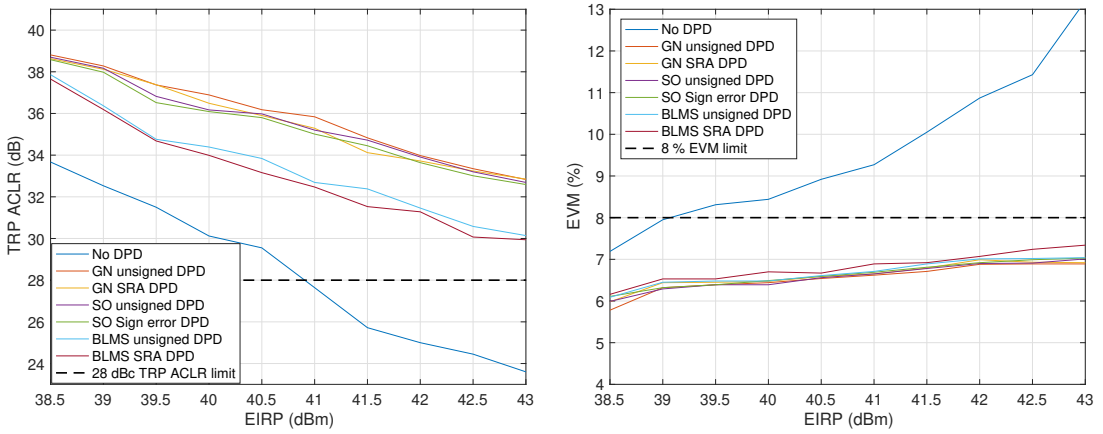
I. Power Sweep with Aggregated 800 MHz Channel BW

The last experiment presents a power sweep similar to that presented in Subsection V-E, but now carried out with the 800 MHz signal. The measured results, illustrated in Fig. 14, are presented again in terms of the TRP ACLR and EVM as functions of the EIRP, sweeping from +38.5 to +43 dBm. Without DPD, the EVM is again the limiting metric at roughly +39 dBm, where both TRP ACLR and EVM requirements are still fulfilled. When applying DPD processing, the EIRP can be further increased to +43 dBm and beyond, as the figure indicates. These results essentially indicate a power efficiency increase of the overall transmitter, since the antenna array can be operated closer to saturation thanks to the DPD, which facilitates increasing the transmit power. The unsigned and signed versions of the studied algorithms are again very close to each other, in terms of performance, yielding similar conclusions as earlier.

In general, the injection-based MP-LUT DPD and the three learning equations and their signed versions studied along this paper demonstrate very appealing performance-complexity trade-offs, while being able to successfully linearize a state-of-the-art active antenna array even with 800 MHz modulation/channel bandwidth. The signed GN and SO can be adopted when a high DPD performance is required, while the signed BLMS approach can be utilized in cases where very small learning complexity is required.

VI. Conclusions

In this article, injection-based memory polynomial LUT DPD system was proposed, together with various signed closed-loop DPD learning algorithms. The described MP-LUT DPD, along with the signed learning algorithms, were



(a) (b)

Fig. 14. Measured TRP-based ACLR in (a), and EVM in (b), as functions of the EIRP in the 800 MHz channel bandwidth case at 28 GHz.

shown to facilitate large reductions in terms of computational complexity, while maintaining a very similar linearization performance compared to ordinary MP DPD. Additionally, the use of LUTs avoids the rank deficiencies in the Signed-Regressor and Sign-sign algorithms, thus eliminating the need for additional matrix transformations that are required in the existing reference methods. Even the block-LMS and its different signed variants were shown to be applicable with the injection-based MP-LUT DPD, without any orthogonalization procedures, allowing for very low complexity. Due to the injection-based DPD structure, it was further shown that the LUT entry sizes required in the DPD system can be decreased, allowing to utilize only 32 or even 16 LUT entries while still achieving efficient linearization performance. Extensive millimeter-wave OTA measurements using a state-of-the-art 28 GHz active antenna array were reported to validate the proposed techniques, incorporating very wide channel bandwidths of 400 MHz and 800 MHz while also pushing the active array very close to saturation. The shown measurement results, together with the detailed complexity analysis, demonstrate that the proposed techniques have a very favorable complexity-performance trade-off in mmW active array linearization.

REFERENCES

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

[2] A. Zhu, "Digital predistortion and its combination with crest factor reduction," in *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*, F.-L. Luo, Ed. Cambridge: Cambridge University Press, 2011, ch. 9, pp. 244–279.

[3] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 4, pp. 1166–1174, April 2012.

[4] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.

[5] C. Mollen, E. G. Larsson, U. Gustavsson, T. Eriksson, and R. W. Heath, "Out-of-band radiation from large antenna arrays," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 196–203, April 2018.

[6] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, 2019.

[7] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital Predistortion for Hybrid MIMO Transmitters," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445–454, June 2018.

[8] A. Brihuega, L. Anttila, M. Abdelaziz, and M. Valkama, "Digital Predistortion in Large-Array Digital Beamforming Transmitters," in *52nd Asilomar Conf. Signals, Syst., and Computers*, Oct. 2018, pp. 611–618.

[9] 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.

[10] X. Liu, Q. Zhang, W. Chen, H. Feng, L. Chen, F. M. Ghannouchi, and Z. Feng, "Beam-Oriented Digital Predistortion for 5G Massive MIMO Hybrid Beamforming Transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 7, pp. 3419–3432, July 2018.

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

[12] 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.

[13] A. Brihuega, M. Abdelaziz, L. Anttila, M. Turunen, M. Allén, T. Eriksson, and M. Valkama, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Tech.*, pp. 1–1, 2020.

[14] 3GPP Tech. Spec. 38.104, "NR; Base Station (BS) radio transmission and reception," v15.4.0 (Release 15), Dec. 2018.

[15] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, p. 1, Nov. 2001.

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

[17] A. Zhu and T. J. Brazil, "An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers," in *2006 IEEE Annual Wireless and Microwave Technology Conference*, Dec. 2006, pp. 1–5.

[18] ——, "Behavioral modeling of RF power amplifiers based on pruned Volterra series," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 12, pp. 563–565, Dec. 2004.

[19] P. P. Campo, V. Lampu, A. Meirhaeghe, J. Boutellier, L. Anttila, and M. Valkama, "Digital predistortion for 5G small cell: GPU implementation and RF measurements," *J. Signal Process. Syst.*, pp. 1–12, 2019.

[20] H. Wang, G. Li, C. Zhou, W. Tao, F. Liu, and A. Zhu, "1-bit observation for direct-learning-based digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2465–2475, Jul. 2017.

[21] N. Guan, N. Wu, and H. Wang, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 921–923, Oct 2018.

[22] Y. Ma, Y. Yamao, Y. Akaiwa, and C. Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 11, pp. 3961–3973, Nov. 2013.

[23] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980–987, Nov. 2017.

[24] L. Ding, R. Raich, and G. Tong Zhou, "A Hammerstein predistortion linearization design based on the indirect learning architecture," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, May 2002, pp. 2689–2692.

[25] H. Huang, J. Xia, and S. Boumaiza, "Novel parallel-processing-based hardware implementation of baseband digital predistorters for linearizing wideband 5G transmitters," *IEEE Trans. Microw. Theory Tech."*, pp. 1–1, 2020.

[26] X. Wu, N. Zheng, X. Yang, J. Shi, and H. Chen, "A spline-based Hammerstein predistortion for 3G power amplifiers with hard nonlinearities," in *2010 2nd International Conference on Future Computer and Communication*, vol. 3, May 2010, pp. V3–741–V3–745.

[27] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 280–293, Jan. 2018.

[28] P. Pascual Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allen, and M. Valkama, "Closed-loop sign algorithms for low-complexity digital predistortion," in *2020 IEEE MTT-S International Microwave Symposium (IMS)*, 2020.

[29] C. D. Presti, D. F. Kimball, and P. M. Asbeck, "Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA PA module," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 3, pp. 604–618, 2012.

[30] R. N. Braithwaite, "General principles and design overview of digital predistortion," in *Digital Front-End in Wireless Communications and Broadcasting*. Cambridge Univ. Press, 2011, ch. 6, pp. 143–191.

[31] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 2001.

[32] M. O'Droma, S. Meza, and Y. Lei, "New modified Saleh models for memoryless nonlinear power amplifier behavioural modelling," *IEEE Commun. Lett.*, vol. 13, no. 6, pp. 399–401, 2009.

[33] Q. Luo, X. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Trans. Microw. Theory Tech.*, vol. 68, no. 1, pp. 301–315, 2020.

[34] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5g massive mimo wireless transmitters based on indirect identification of power amplifier behavior with ota tests," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 1, pp. 316–328, 2020.

[35] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 12, pp. 5235–5248, 2019.

[36] F. Jalili, F. F. Tafuri, O. K. Jensen, Y. Li, M. Shen, and G. F. Pedersen, "Linearization trade-offs in a 5G mmWave active phased array OTA setup," *IEEE Access*, vol. 8, pp. 110 669–110 677, 2020.

[37] L. Guan and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 4, pp. 866–872, Apr. 2010.

[38] J. A. Tropp, "Recovery of short, complex linear combinations via l1 minimization," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1568–1570, Apr. 2005.

[39] N. Guan, N. Wu, and H. Wang, "Digital predistortion of wideband power amplifier with single undersampling ADC," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 11, pp. 1016–1018, Nov 2017.

[40] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3, pp. 238–239.

[41] Lei Ding, G. T. Zhou, D. R. Morgan, Zhengxiang Ma, J. S. Kenney, Jaehyeong Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, 2004.

Pablo Pascual Campo received his B.Sc. and M.Sc. degrees in Telecommunications and Electrical Engineering in 2016 and 2018, respectively, from Universidad Politécnica de Madrid, Madrid, Spain. He is currently pursuing his D.Sc. degree at Tampere University, Department of Electrical Engineering, Tampere, Finland. His research interests include digital predistortion, full-duplex systems and applications, and signal processing for wireless communications at the mmWave bands.

Vesa Lampu received the B.Sc. and M.Sc. degrees in electrical engineering in 2017 and 2019, respectively from Tampere University (formerly TUT), Tampere, Finland, where he is currently working toward the D.Sc. degree. He has been with the Department of Electrical Engineering, Tampere University since 2018. His research focuses on energy-efficient multiple-input– multiple-output systems. He has also been involved in research on nonlinear system identification, with applications in full-duplex and digital predistortion techniques.

Lauri Anttila received the D.Sc. (Tech.) degree (with distinction) in 2011 from Tampere University of Technology (TUT), Finland. Since 2016, he has been a University Researcher at the Department of Electrical Engineering, Tampere University (formerly TUT). In 2016-2017, he was a Visiting Research Fellow at Aalto University, Finland. His research interests are in radio communications and signal processing, with a focus on the radio implementation challenges in systems such as 5G, full-duplex radio, and large-scale antenna systems.

Alberto Brihuega (S'18) received the B.Sc. and M.Sc. degrees in Telecommunications Engineering from Universidad Politécnica de Madrid, Spain, in 2015 and 2017, respectively. He is currently working towards the Ph.D. degree with Tampere University, Finland, where he is a researcher with the Department of Electrical Engineering. His research interests include statistical and adaptive digital signal processing for compensation of hardware impairments in large-array antenna transceivers.

Markus Allén received the B.Sc., M.Sc. and D.Sc. degrees in communications engineering from Tampere University of Technology, Finland, in 2008, 2010 and 2015, respectively. He is currently with the Department of Electrical Engineering at Tampere University, Finland, as a University Instructor. His current research interests include software-defined radios, 5G-related RF measurements and digital signal processing for radio transceiver linearization.

Yan Guo received the B.E. degree in information science and engineering from East China Jiaotong University, Nanchang, Jiangxi Province, China, in 2007, the M.E. degree in communication and information systems from Southeast University, Nanjing, China, in 2011, and the Ph.D degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2016. Dr.Guo is now a principle research engineer and a research team leader with Wireless Terminal Chipset Algorithm Department, of Hisilicon, Huawei. His main research interesting including nonlinear behavioral modeling and digital linearization techniques for RF front-end of broadband UE.

Mikko Valkama received the D.Sc. (Tech.) degree (with honors) from Tampere University of Technology (TUT), Finland, in 2001. In 2003, he was a visiting post-doc research fellow with SDSU, San Diego, CA. Currently, he is a Full Professor and Department Head of Electrical Engineering at the newly formed Tampere University (TAU), Finland. His research interests include radio communications, radio localization, and radio-based sensing, with emphasis on 5G and beyond mobile radio networks.

PUBLICATION 3

Cascaded spline-based models for complex nonlinear systems: methods and applications

P. Pascual Campo, L. Anttila, D. Korpi and M. Valkama

IEEE Transactions on Signal Processing 69.(2021), 370–384
DOI: 10.1109/TSP.2020.3046355

Publication reprinted with the permission of the copyright holders

Cascaded Spline-Based Models for Complex Nonlinear Systems: Methods and Applications

Pablo Pascual Campo, *Graduate Student Member, IEEE*, Lauri Anttila, *Member, IEEE*, Dani Korpi, and Mikko Valkama, *Senior Member, IEEE*

***Abstract*—In this paper, we present a class of cascaded nonlinear models for complex-valued system identification, aimed at baseband modeling of nonlinear radio systems. The proposed models consist of serially connected elementary linear and nonlinear blocks, with the nonlinear blocks implemented as uniform spline-interpolated look-up tables (LUT) and the linear blocks as FIR filters. Wiener, Hammerstein, and Wiener-Hammerstein models are built, and simple but efficient gradient based adaptive learning rules are derived for all the models. This approach leads to remarkably simple solutions in terms of computational complexity, making the techniques suitable for real-time implementation. The proposed methods are then applied to full-duplex self-interference cancellation and digital predistortion in various real-life scenarios. First, evaluations with measured data from an in-band full-duplex prototype working at 2.4 GHz ISM band show that the algorithms are capable of obtaining similar cancellation performance as existing state-of-the-art solutions, regardless of the clearly reduced complexity. Second, a mmW active antenna array working at 28 GHz center frequency is digitally predistorted with the proposed solutions. The unwanted emissions and nonlinear distortion are suppressed to similar levels as with other state-of-the art solutions, and the corresponding linearity specifications are fulfilled in all cases, while the processing complexity is again drastically reduced.**

***Index Terms*—Behavioral modeling, cascaded models, digital predistortion, full-duplex, linearization, look-up tables, power amplifiers, self-interference cancellation, splines.**

I. Introduction

NONLINEAR modeling and system identification are important ingredients in many wireless communications systems, due to the inherent nonlinearity of certain hardware components. The power amplifier (PA), for example, is typically operated in its nonlinear region to improve power efficiency [1]. Behavioral modeling and digital predistortion (DPD) of the PA are typical examples of nonlinear modeling tasks [2]. Other applications include, for example, digital self-interference cancellation (DSIC) in simultaneous transmit and receive (STAR) devices [3]–[5] and nonlinear channel equalization in satellite communications [6], [7].

Most often a general black-box or behavioral modeling approach is adopted for such problems, for example based on the Volterra series [8], [9] or artificial neural networks [7], [10]. However, in many cases some physical knowledge of the system under study is available, and a block-oriented model is well argued [7]. A good example of such a problem is self-interference cancellation in STAR or in-band full-duplex (IBFD), where the modeling is typically concerned with identifying a nonlinear device (the PA) in cascade with a linear system (the self-interference propagation channel), which together constitute a Hammerstein system [11]. Nonetheless, cascaded models have been successfully applied to black-box modeling problems as well such as DPD, see for example [12]–[15]. Compared to linear-in-parameters models such as Volterra, cascaded models typically have much less free parameters, thus they are appealing when low-complexity implementation is pursued.

The literature is abundant in complex nonlinear models and learning techniques for PA behavioral modeling, DPD, and DSIC. The vast majority of the existing works are based on Volterra-type models; see [2], [16] for overviews of these approaches. The literature on cascaded models and the associated learning methods for communications applications is more scarce. Techniques in this category contain the works in [11]–[15], [17]–[22]. In [20], [21], a frequency-domain identification of Hammerstein and Wiener-Hammerstein DPD models, respectively, was proposed. These works relied on the direct learning approach, i.e., required learning the PA forward model which was then inverted. In [12], a Hammerstein DPD with polynomial nonlinearity was proposed, together with an offline two-stage least-squares (LS) estimation scheme based on the indirect learning architecture (ILA). In [18], the authors proposed a Hammerstein DPD with a look-up table (LUT) nonlinearity. They also resorted to a two-stage estimation procedure and ILA, utilizing smoothed AM-AM and AM-PM curves for generating the LUT, and LS for estimating the filter coefficients. The study in [14] proposed a parallel Hammerstein (PH) DPD, with two independent real-valued spline-based Hammerstein models for the magnitude and phase. ILA with the Levenberg-Marquardt algorithm was used for parameter estimation, making real-time learning and tracking challenging due to the involved computational complexity. In [17] and [15], the authors presented Wiener and Hammerstein models, respectively, building on spline-based

Manuscript received June 23, 2020; revised November 11, 2020; accepted December 12, 2020. Date of publication December 23, 2020; date of current version January 14, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. A. A. Rontogiannis. This work was supported in part by the Academy of Finland under the projects 301820, 323461, 332361, and 319994. *(Corresponding author: Pablo Pascual Campo.)*

Pablo Pascual Campo, Lauri Anttila, and Mikko Valkama are with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland (e-mail: pablo.pascualcampo@tuni.fi; lauri.anttila@tuni.fi; mikko.valkama@tuni.fi).

Dani Korpi is with Nokia Bell Labs, 02610 Espoo, Finland (e-mail: dani.korpi@nokia-bell-labs.com).

Digital Object Identifier 10.1109/TSP.2020.3046355

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

neural networks. The learning is based on the direct learning architecture, where the forward model is estimated with LS based algorithms, and then inverted with an iterative scheme. Altogether, the learning complexities of all the above techniques are considerable, and can be considered generally unsuitable for real-time implementation. In [22], a Hammerstein DPD based on the simplical canonical piece-wise linear (SCPWL) basis functions was introduced, and least mean squares (LMS)-based learning algorithms with both direct and indirect learning architecture were proposed. However, the solutions required the identification of the forward PA model, which was assumed to obey the Wiener model, thus complicating the estimation procedure. The Hammerstein DPD model was also formulated into a linear-in-parameters form (effectively a PH model), which led to overparameterization.

In the context of DSIC, PH based algorithms building on polynomial nonlinearities were proposed in [23]–[26], with [23]–[25] operating in time-domain and [26] in the frequency-domain. The works in [23], [24] utilized block LS estimation, while adaptive estimation methods were used in [25], [26]. While the PH model has been shown to be an accurate model for DSIC, it suffers from high complexity, making real-time implementation very challenging, especially with wider bandwidths. In [11], [27], which serve as a starting point or background for this paper, we proposed a Hammerstein model for DSIC and DPD, respectively, building on spline-interpolated LUTs and simple adaptive learning rules.

In this paper, we propose block-oriented models for complex-valued baseband modeling of nonlinear radio systems, along with simple gradient-based learning rules. We build Wiener, Hammerstein, and Wiener-Hammerstein models, comprised of different combinations of spline-interpolated LUTs and finite impulse response (FIR) filters. The techniques are based on the recently introduced concept of spline adaptive filters, which was developed for real-valued systems in [28]–[30]. The proposed techniques differ from these works in the following ways:

- The adaptive spline interpolation scheme within the cascaded models is specifically developed for complex-valued modeling of radio systems, instead of real-valued systems;
- We introduce a so-called injection-based nonlinear block, where the complex nonlinear gain of the system is defined as a deviation from unity gain. This helps to control the gain ambiguity between the cascaded blocks, and also reduces the dynamic range and thus the number of bits of the LUT;
- We point out and fix a shortcoming in the original learning equations in [28]–[30], which slowed down their convergence;
- We apply the techniques to two real-life problems: self-interference cancellation in IBFD, and digital predistortion of nonlinear power amplifiers;
- We verify the functionality and performance of the models using measured signals, and compare to the widely-used memory polynomial (MP) and generalized memory polynomial (GMP).

The proposed techniques are shown to offer similar performance compared to the MP and GMP models in both direct and inverse modeling problems using measured signals, but with greatly reduced processing complexity. Altogether, the framework proposed in this paper offers appealing low-complexity adaptive solutions for real-time applications that require complex-valued nonlinear model identification. Thus the paper offers contributions in both signal processing theory and applications.

The rest of the paper is organized as follows. First, Section II presents the general real-valued theory of spline interpolation, and then extends it to the complex domain. Building on our early work in [11], Section III then presents the spline-interpolated Hammerstein model, provides a complexity analysis, and highlights the differences to the original real-valued work in [29]. Section IV and Section V describe the spline-interpolated Wiener and Wiener-Hammerstein approaches, along with their learning rule derivations and complexity analyses. In Section VI, we present two applications for the techniques, along with experiments in real-life use cases, in order to verify and validate the proposed models. Finally, Section VII summarizes the main findings of this paper.

I. Notation used in this paper

In this paper, matrices are represented by capital boldface letters, e.g., $\mathbf{\Sigma} \in \mathbb{C}^{M\times N}$. Ordinary transpose, Hermitian transpose, and complex conjugation are denoted by $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. By default, vectors are complex-valued column vectors, presented with lowercase boldface letters, i.e., $\mathbf{v} \in \mathbb{C}^{M\times 1} = [v_1\, v_2\, \cdots\, v_M]^T$. Additionally, the absolute value, floor, and ceil operators are represented as $|\cdot|$, $\lfloor\cdot\rfloor\cdot$, and $\lceil\cdot\rceil\cdot$, respectively.

II. Principles of B-Spline Interpolation

This section presents the B-spline interpolation theory. First, the traditionally used real-valued scheme is presented, and second, our extension to the complex domain is introduced. This will enable the use of the scheme in the context of radio communications, where complex I/Q signals are utilized.

A. Real-Valued B-Spline Interpolation

Spline interpolation builds on piece-wise polynomials to interpolate between an arbitrary set of points known as control points, under certain continuity and smoothness constraints at the connecting points. With such a piece-wise modeling approach, simpler and lower-order functions can be adopted per individual region, in contrast to the classical polynomial methods where a single high-order expression is utilized to model the whole input range. The use of lower-order functions essentially translates to reduced associated processing complexities [11], [29], as demonstrated in later sections. Also, it allows to better condition the estimation problem [31], i.e., obtaining a lower condition number for the regression matrix and thus avoiding the need of prewhitening/orthogonalization. Another potential advantage of piece-wise models is the ability to use different polynomial orders for different regions, depending on the desired modeling accuracy in each.

To build the piece-wise scheme, we formally define a set of *knots* that divides the input data range into N regions, defined as $\mathbf{T} = \{t_0, t_1, \ldots, t_N\}$, corresponding to regions $i = \{1, 2, \ldots, N\}$. The *knots* are constrained to be non-decreasing, i.e., $t_0 < t_1 < \cdots < t_N$. We consider uniform splines in this paper, since it allows for a simple input-output relation, suitable for adaptive estimation of the control points. Thus, the region width is defined as $\Delta = t_{i+1} - t_i,\ \Delta > 0$.

In general, this construction can be generalized to any number of regions, N, and any spline degree, P. The B-spline segment describing each individual region, $[t_i, t_{i+1})$, is therefore an affine combination of $P+1$ spline curves [32]. Each of these curves is characterized by the Pth-degree spline basis function, given by the De Boor recursion [31] as

$$N_i^P(u) = \frac{u - t_i}{t_{i+P} - t_i} N_i^{P-1}(u) + \frac{t_{i+P+1} - u}{t_{i+P+1} - t_{i+1}} N_{i+1}^{P-1}(u), \tag{2}$$

where the initial 0th-order basis function, $N_i^0(u)$, can be defined as

$$N_i^0(u) = \begin{cases} 1 \text{ if } t_i \le u < t_{i+1}, \\ 0 \text{ otherwise.} \end{cases} \tag{3}$$

The nonzero portions of these segments span over the interval $[t_i, t_{i+P}]$. Additionally, the blending functions N_i^P for different regions are shifted versions of each other, and in general, they can be written as [33]

$$N_i^P(u) = N_0^P(u - i). \tag{4}$$

Consequently, the resulting spline segment in a particular region can be defined as [34]

$$\gamma(u, i) = \sum_{n=i-P-1}^{i-1} N_n^P(u) q_n, \tag{5}$$

where q_n is the corresponding control point multiplying each spline curve. Alternatively, equation in (5) can be expressed as the inner product of two vectors, as

$$\gamma(u,i) = \begin{bmatrix} N_{i-P-1}^P(u) & N_{i-P}^P(u) & \cdots & N_{i-1}^P(u) \end{bmatrix} \begin{bmatrix} q_{i-P-1} \\ q_{i-P} \\ \vdots \\ q_{i-1} \end{bmatrix}. \tag{6}$$

Further developing this expression by substituting the recursion in (2) [28], and combining the spline segments of all the regions, a generic formulation as a function of the spline basis functions is obtained as

$$\gamma(u, i) = \mathbf{\Psi}^T \mathbf{q}, \tag{7}$$

where $\mathbf{q} \in \mathbb{R}^{Q \times 1} = [q_0 \quad q_1 \quad \cdots \quad q_Q]^T$ is the complete set of spline control points. We note that the total number of control points with N regions and spline order P is $Q = N + P$. Additionally, $\mathbf{\Psi}$ reads

$$\mathbf{\Psi} \in \mathbb{R}^{Q \times 1} = [0 \ \cdots \ 0 \ \mathbf{u}^T \mathbf{C}_P \ 0 \ \cdots \ 0]^T. \tag{8}$$

Here, $\mathbf{u}$ is referred to as the abscissa vector, defined as

$$\mathbf{u} \in \mathbb{R}^{(P+1) \times 1} = [u^P \quad u^{P-1} \quad \cdots \quad 1]^T, \tag{9}$$

and $\mathbf{C}_P \in \mathbb{R}^{(P+1) \times (P+1)}$, shown at the bottom of this page, in (1) for $P = 1, \cdots, 4$, is the spline basis matrix, which depends on the chosen spline order P and the knot spacing Δ. The term $\mathbf{u}^T \mathbf{C}_P$ in $\mathbf{\Psi}$ is indexed such that the starting index is i (i.e. $i-1$ leading zeros and $Q - (i + P)$ trailing zeros in (8)), such that only the corresponding control points are contributing to the basis function weighting, according to the region index.

B. Complex-Valued B-Spline Interpolation

In the context of radio communications, complex-valued I/Q signals are utilized. Therefore, real-valued spline interpolation theory needs to be extended to the complex domain. We begin by noting that a memoryless baseband model of a bandpass nonlinear device (such as a PA), with input signal $x[n]$, can be expressed as [35]

$$y[n] = x[n] G(|x[n]|), \tag{10}$$

where $G(|x[n]|) = G_{\mathrm{I}}(|x[n]|) + jG_{\mathrm{Q}}(|x[n]|)$ is the nonlinear complex gain of the device. Notice that $G(|x[n]|)$ depends only on the magnitude of the input signal, and not on its phase.

Let us now denote $x[n]$ as the input signal and $G(|x[n]|)$ as the output of the spline nonlinearity. According to (10), two separate splines can be used to model the I and Q responses. As the input to the splines is a unipolar magnitude signal, the region index i_n and abscissa value u_n, at time instant n, are defined as

$$i_n = \left\lfloor \frac{|x[n]|}{\Delta_{\mathrm{x}}} \right\rfloor + 1, \tag{11}$$

$$u_n = \frac{|x[n]|}{\Delta_{\mathrm{x}}} - (i_n - 1), \tag{12}$$

$$\mathbf{C}_1 = \begin{pmatrix} \frac{-1}{\Delta} & \frac{1}{\Delta} \\ 1 & 0 \end{pmatrix}, \ \mathbf{C}_2 = \frac{1}{2} \begin{pmatrix} \frac{1}{\Delta^2} & \frac{-1}{\Delta^2} & \frac{1}{\Delta^2} \\ \frac{-2}{\Delta} & \frac{2}{\Delta} & 0 \\ 1 & 1 & 0 \end{pmatrix}, \ \mathbf{C}_3 = \frac{1}{6} \begin{pmatrix} \frac{-1}{\Delta^3} & \frac{3}{\Delta^3} & \frac{-3}{\Delta^3} & \frac{1}{\Delta^3} \\ \frac{3}{\Delta^2} & \frac{-6}{\Delta^2} & \frac{3}{\Delta^2} & 0 \\ \frac{-3}{\Delta} & 0 & \frac{3}{\Delta} & 0 \\ 1 & 4 & 1 & 0 \end{pmatrix}, \ \mathbf{C}_4 = \frac{1}{24} \begin{pmatrix} \frac{1}{\Delta^4} & \frac{-4}{\Delta^4} & \frac{6}{\Delta^4} & \frac{-4}{\Delta^4} & \frac{1}{\Delta^4} \\ \frac{-4}{\Delta^3} & \frac{12}{\Delta^3} & \frac{-12}{\Delta^3} & \frac{4}{\Delta^3} & 0 \\ \frac{6}{\Delta^2} & \frac{-6}{\Delta^2} & \frac{-6}{\Delta^2} & \frac{6}{\Delta^2} & 0 \\ \frac{-4}{\Delta} & \frac{-12}{\Delta} & \frac{4}{\Delta} & \frac{12}{\Delta} & 0 \\ 1 & 11 & 11 & 1 & 0 \end{pmatrix}. \tag{1}$$

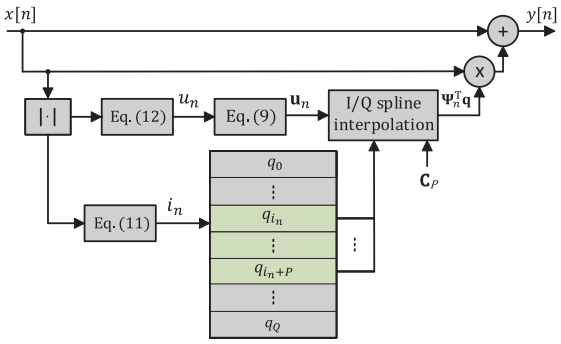


Fig. 1. The injection-based complex spline interpolated LUT scheme presented in (16), utilized in the proposed block-oriented models.

where Δ_{x} is the region width, and u_n is within the interval $[0, \Delta_{\mathrm{x}})$.

The outputs of the two splines modeling the I and Q branch responses of the system can then be written as

$$G_{\mathrm{I}}(|x[n]|) = \mathbf{\Psi}_n^T \mathbf{q}^{\mathrm{re}}, \tag{13}$$

$$G_{\mathrm{Q}}(|x[n]|) = \mathbf{\Psi}_n^T \mathbf{q}^{\mathrm{im}}, \tag{14}$$

where $\mathbf{q}^{\mathrm{re}}$ and $\mathbf{q}^{\mathrm{im}}$ contain the control points for the I and Q splines, and $\mathbf{\Psi}_n$ is defined in (8). Following (10), and using (13) and (14), the spline model output can be written as

$$\begin{aligned} y[n] &= x[n](G_{\mathrm{I}}(|x[n]|) + jG_{\mathrm{Q}}(|x[n]|)) \\ &= x[n]\mathbf{\Psi}_n^T(\mathbf{q}^{\mathrm{re}} + j\mathbf{q}^{\mathrm{im}}) \\ &= x[n]\mathbf{\Psi}_n^T\mathbf{q}, \end{aligned} \tag{15}$$

where $\mathbf{q} = \mathbf{q}^{\mathrm{re}} + j\mathbf{q}^{\mathrm{im}}$ now contains the complex-valued control points.

In this paper, we define the complex nonlinear gain function as a deviation from unity gain, and thus write $G(|x[n]|) = 1 + G_{\mathrm{I}}(|x[n]|) + jG_{\mathrm{Q}}(|x[n]|)$. Thus, with this assumption, the output of the nonlinear subsystem is written as

$$\begin{aligned} y[n] &= x[n] + x[n]\mathbf{\Psi}_n^T\mathbf{q} \\ &= x[n]\mathbf{\Psi}_n^T(\mathbf{1} + \mathbf{q}), \end{aligned} \tag{16}$$

and further illustrated in Fig. 1. Here, $\mathbf{1} \in \mathbb{R}^{Q\times 1}$ denotes a vector of all ones, and the partition of unity property of B-splines [33], or $\mathbf{\Psi}_n^T\mathbf{1} = 1$, is used to arrive to the last form. Applying such a nonlinear subsystem in cascaded models will effectively remove the gain ambiguity between the linear and nonlinear blocks. It will also reduce the dynamic range of the control point vector $\mathbf{q}$, thus requiring less bits in a fixed-point implementation. This will be illustrated in the following subsection.

C. Dynamic Range of $\mathbf{q}$

To illustrate the reduced dynamic range, the Modified Saleh (MS) memoryless PA model presented in [36] is approximated with the spline models drawn in (15) and (16). The MS AM-AM and AM-PM responses can be described by

$$z(r) = \frac{\alpha_z r}{\sqrt{1+\beta_z r^3}}; \quad \phi(r) = \frac{\alpha_\phi}{\sqrt[3]{1+\beta_\phi r^4}} - \epsilon; \tag{17}$$

where r and z represent the magnitudes of the input and output signal, respectively, and ϕ is the phase error of the output signal. Additionally, α_z, β_z, α_ϕ, β_ϕ and ϵ are the AM-AM and AM-PM model coefficients. Herein, it is considered $\alpha_z = 0.82$, $\beta_z = 0.29$, $\alpha_\phi = -0.35$, $\beta_\phi = 1$ and $\epsilon = -0.36$ [36], which were extracted from PA input/output data measured from an LDMOS PA.

An arbitrary OFDM input signal is then passed through the MS model, and a LS algorithm is used to fit both spline techniques to the output model data, thus obtaining the control point vectors from (15) and (16). The magnitude and phase of these values are then quantized with the same fixed number of bits to illustrate the dynamic range reduction. Fig. 2(a) shows the AM-AM response of the MS and the fitted spline models, with both magnitude and phase quantized with 12 bits. It is clearly seen that the modelling accuracy of (15) is reduced when considering the same number of bits as in (16). Additionally, Fig. 2(b) shows the NMSE values of both solutions as the control point vectors are quantized with an increasing number of bits. The modelling performance of (16) is consistently about 18 dB better than that of (15), until the curves essentially saturate. The exact difference depends on the shape of the nonlinear response. However, assessing this further is out of the scope of this paper.

III. Complex Spline-Based Adaptive Hammerstein Model

The adaptive Hammerstein solution aims at identifying an unknown nonlinear system consisting of a memoryless nonlinear function followed by a linear FIR filter [29]. In our previous publication [11], we studied this structure and applied the spline interpolated LUT as the nonlinear block. This model, herein referred to as SPH, serves as a starting point for this paper, and thus we recapitulate the main expressions.

Let us denote by $x[n]$ and $y[n]$ the input and output signals of the SPH model, and by $s[n]$ the intermediate signal after the nonlinear function. Following the B-spline interpolation procedure in (16) for the nonlinear block, and a classical convolution for the FIR filter, the intermediate and output signals of the SPH model can be written as

$$\begin{aligned} s[n] &= x[n] + x[n]\mathbf{\Psi}_n^T\mathbf{q}_n, \\ y[n] &= \mathbf{g}_n^T\mathbf{s}_n, \end{aligned} \tag{18}$$

where $\mathbf{g}_n \in \mathbb{C}^{M_{\mathrm{g}}\times 1} = [g[0] \quad g[1] \ \cdots \ g[M_{\mathrm{g}}-1]]^T$, $\mathbf{s}_n \in \mathbb{C}^{M_{\mathrm{g}}\times 1} = [s[n]\, s[n-1] \ \cdots \ s[n-M_{\mathrm{g}}+1]]^T$, and M_{g} denotes the number of taps of the linear filter.

A. Learning Rules and Complexity Analysis

The learning rules for both $\mathbf{g}_n$ and $\mathbf{q}_n$ can be obtained by following the gradient-descent approach. From [11], these expressions read

$$\mathbf{g}_{n+1} = \mathbf{g}_n + \mu_{\mathrm{g}} e[n]\mathbf{s}_n^*, \tag{19}$$

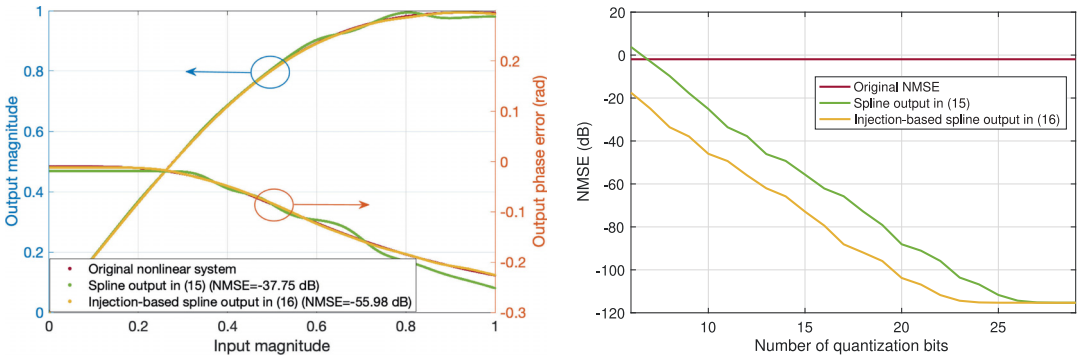


Fig. 2. Response of the Modified Saleh memoryless nonlinear model and its modelling with (15) and (16) and with 12 quantization bits in (a), and NMSE as a function of the LUT quantization bits in (b).

TABLE I
NUMBER OF REQUIRED ARITHMETIC OPERATIONS IN EACH ITERATION OF THE SPH IDENTIFICATION ALGORITHM, IN TERMS OF FLOPS PER SAMPLE, AND REAL MULTIPLICATIONS PER SAMPLE

Operation		FLOPs/sample	Real multiplications/sample
Model identification	$s[n]$	$2P^2+6P+18$	$P^2+4P+8+sqrt$
	$y[n]$	$8M_{\mathrm{g}}-2$	$4M_{\mathrm{g}}$
	Total	$2P^2+6P+8M_{\mathrm{g}}+16$	$P^2+4P+4M_{\mathrm{g}}+8+sqrt$
Coefficient updates	$\mathbf{g}_{n+1}$	$8M_{\mathrm{g}}+2$	$4M_{\mathrm{g}}+2$
	$\mathbf{q}_{n+1}$	$4P\tau+4P+10\tau+3Q+6$	$4P+2P\tau+6\tau+6$
	Total	$4P\tau+4P+10\tau+8M_{\mathrm{g}}+3Q+8$	$4P+2P\tau+6\tau+4M_{\mathrm{g}}+8$
Total		$2P^2+10P+4P\tau+16M_{\mathrm{g}}+10\tau+3Q+24$	$P^2+8P+2P\tau+6\tau+8M_{\mathrm{g}}+16+sqrt$

$$\mathbf{q}_{n+1}=\mathbf{q}_n+\mu_{\mathrm{q}}e[n]\,\boldsymbol{\Xi}_n^T\mathbf{X}_n^*\mathbf{g}_n^*, \tag{20}$$

where $\mathbf{X}_n$ contains the signal regression of $x[n]$ over the span of $\mathbf{g}_n$ in its main diagonal, $\boldsymbol{\Xi}_n\in\mathbb{R}^{M_{\mathrm{g}}\times Q}=[\boldsymbol{\Psi}_n\ \boldsymbol{\Psi}_{n-1}\ \cdots\ \boldsymbol{\Psi}_{n-M_{\mathrm{g}}+1}]^T$, and $e[n]=d[n]-y[n]$, $d[n]$ being the observed signal. Additionally, in the learning rule for $\mathbf{q}_{n+1}$, it is assumed that the rate of change of $\mathbf{q}_n$ over the span of the filter length M_{g} is negligible, i.e., $\mathbf{q}_n\approx\mathbf{q}_{n+M_{\mathrm{g}}}$. This is a plausible assumption since μ_{q} is small.

It is important to notice the relatively high complexity involved in calculating the term $\boldsymbol{\Xi}_n^T\mathbf{X}_n^*\mathbf{g}_n^*$ in the update of $\mathbf{q}_{n+1}$, as $\boldsymbol{\Xi}_n^T$ is a $Q\times M_{\mathrm{g}}$ matrix, somewhat large if $\mathbf{g}_n$ contains a large number of taps. To ease this update, in [11] we proposed a complexity reduction approximation where only a specific temporal span of the matrix $\boldsymbol{\Xi}_n$ (row dimension) is chosen for the update. Obviously, the largest filter taps in $\mathbf{g}_n$ are the ones that contribute the most to the update, and those taps typically correspond to the nearest past samples. Thus, only τ taps, i,e,. rows of $\boldsymbol{\Xi}_n$, $\mathbf{g}_n$, and $\mathbf{x}_n$, can be selected for the learning rule, simplifying the overall update and having little effect in the final modeling performance. This is

$$\boldsymbol{\Xi}_n\in\mathbb{R}^{\tau\times Q}=\left[\boldsymbol{\Psi}_n\ \ \boldsymbol{\Psi}_{n-1}\ \ \cdots\ \ \boldsymbol{\Psi}_{n-\tau+1}\right]^T. \tag{21}$$

We refer the reader to [11] for a more detailed presentation of the SPH model.

Finally, we analyze the computational complexity of the SPH model in terms of FLOPs and real multiplications per sample. The results are gathered in Table I. For this complexity analysis, it is assumed that one complex multiplication is calculated with 6 FLOPs or 4 real multiplications, while one complex-real multiplication costs 2 FLOPs or 2 real multiplications, and one real addition costs 1 FLOP and 0 real multiplications. Additionally, one square root costs 8 FLOPs, and it is left indicated as $sqrt$ in the case of real multiplications [16]. Equivalently, one division is left indicated as div in the complexity tables.

We also note that in certain applications, it may not be necessary to update all coefficients at every iteration. One particular example is in the full-duplex scenario, shown in Section VI-A, where the control points $\mathbf{q}_n$ can remain static once steady-state is reached, as the operating conditions of the PA usually do not change drastically once estimated. The filter coefficients, on the other hand, require constant updating, because the self-interference (SI) channel is generally time-varying.

B. Suboptimal Formulation in Scarpiniti Et Al.

As mentioned in the introduction to this paper, the original real-valued presentation of the techniques contained a defect in

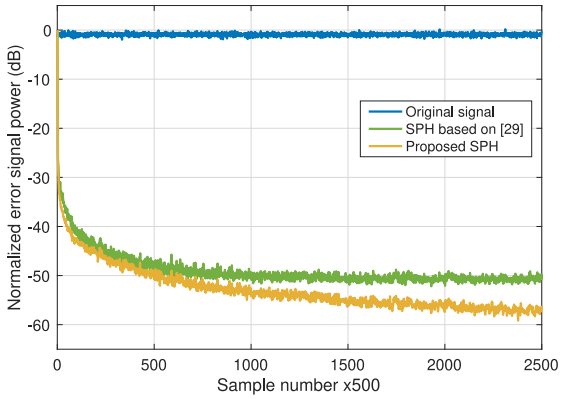


Fig. 3. Convergence comparison between the SPH algorithm based on [29] and the proposed SPH algorithm.

the coefficient updates, which affected the convergence. In [29], the update of the real-valued control point vector $\mathbf{q}_n$ was defined as

$$\mathbf{q}_{i,n+1} = \mathbf{q}_{i,n} + \mu_{\mathrm{q}} e[n]\, \mathbf{C}_P^T \mathbf{U}_{i,n} \mathbf{g}_n. \tag{22}$$

Here, $\mathbf{q}_{i,n} = [q_{i,n}\ q_{i+1,n}\ \cdots\ q_{i+P,n}]^T$ with $q_{k,n}$ denoting the k-th entry of $\mathbf{q}_n$, is the subset of $\mathbf{q}_n$ starting from index i. The matrix $\mathbf{U}_{i,n} = [\mathbf{u}_{i,n}\mathbf{u}_{i,n-1}\ldots\mathbf{u}_{i,n-M+1}]$ collects M past vectors $\mathbf{u}_{i,n-k}$, where each vector assumes the value of $\mathbf{u}_{n-k}$ if evaluated in the same span as the current input sample or in an overlapped span, while otherwise it is a zero vector. The equation in (22) thus only updates the subset $\mathbf{q}_{i,n}$ at time instant n, whereas our correct learning rule in (20) updates the whole vector $\mathbf{q}_n$ in each iteration. The update rule in (22) can also be seen to use some samples from wrong spans to update the current span $\mathbf{q}_{i,n}$, therefore making it suboptimal. To substantiate this, we derived the complex-valued version of (22), which reads

$$\mathbf{q}_{i,n+1} = \mathbf{q}_{i,n} + \mu_{\mathrm{q}} e[n]\, \mathbf{C}_P^T \mathbf{U}_{i,n} \mathbf{X}_n^* \mathbf{g}_n^*, \tag{23}$$

and compared its learning behavior with the update in (20), when learning a cascaded Hammerstein system. Specifically, the Hammerstein structure contains the Saleh nonlinear function presented in Section II-C, followed by an FIR filter with 20 complex-valued taps to account for the memory effects. The same parametrization ($P=3$, $M=20$, $Q=7$, $\mu_{\mathrm{w}}=0.002$, $\mu_{\mathrm{q}}=0.002$) is naturally chosen in both models to ensure fairness, and the resulting error signal powers are then presented in Fig. 3. The proposed Hammerstein solution is seen to clearly outperform the one based on [29] in terms of convergence speed and final steady-state behavior.

We finally note that, although the whole control point vector in (20) is updated in every iteration, only the control points corresponding to the current span (i.e. $q_{i_n}, q_{i_n+1}, \ldots, q_{i_n+P}$) are selected to build the output signal, $y[n]$, in (16). This stems from the definition of $\mathbf{\Psi}_n$, which indexes the term $\mathbf{u}^T\mathbf{C}_P$ starting from index i_n, such that only the correct control points are selected for the interpolation.

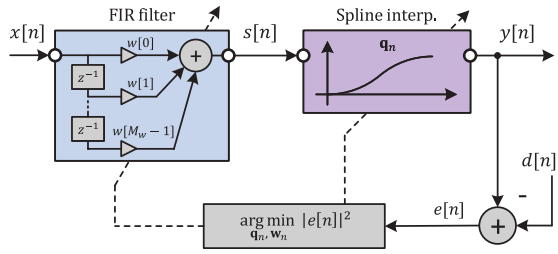


Fig. 4. Architecture of the complex spline interpolated Wiener system. Note that the *spline interpolation* block incorporates the scheme presented in Fig. 1.

IV. Complex Spline-Based Adaptive Wiener Model

In this section, we present the details of the adaptive Wiener solution, herein called SPW, which builds on spline interpolated LUTs. This technique, outlined in Fig. 4, models an unknown nonlinear system with memory effects by cascading a linear filter and a nonlinear function. We denote the linear filter by $\mathbf{w}$, and its tap number by $M_{\mathrm{w}} = M_{\mathrm{w,pre}} + M_{\mathrm{w,post}} + 1$, considering both pre-cursor and post-cursor taps. For simplicity, the model equations consider $M_{\mathrm{w,pre}} = 0$. This consideration is also taken in other following methods.

First, denoting $x[n]$ as the input signal of the model, the intermediate signal $s[n]$ after the linear filter can be written as

$$s[n] = \mathbf{w}_n^T \mathbf{x}_n, \tag{24}$$

where $\mathbf{w}_n \in \mathbb{C}^{M_{\mathrm{w}}\times 1} = [w[0]\ \ w[1]\ \cdots\ w[M_{\mathrm{w}}-1]]^T$, and $\mathbf{x}_n \in \mathbb{C}^{M_{\mathrm{w}}\times 1} = [x[n]\ x[n-1]\ \cdots\ x[n-M_{\mathrm{w}}+1]]^T$.

Secondly, the nonlinear subsystem in (16) is excited with the filter output signal $s[n]$, yielding the SPW model output

$$y[n] = s[n] + s[n]\mathbf{\Psi}_n^T \mathbf{q}_n. \tag{25}$$

A. Learning Rules

In order to make the Wiener model adaptive, two different learning rules, estimating the filter coefficients $\mathbf{w}_n$ and control points $\mathbf{q}_n$, are derived. To this end, we define the error signal

$$e[n] = d[n] - y[n], \tag{26}$$

where the model output signal $y[n]$ is subtracted from the observed signal, or the desired response of the system, denoted here by $d[n]$. The problem lies now in estimating the values of $\mathbf{w}_n$ and $\mathbf{q}_n$ to minimize $e[n]$. This can be done by following the classical gradient-descent solution, where the quantities are learned by following the negative direction of the gradient, towards a local minimum in the mean surface error [37]. The coefficients can be updated iteratively when necessary, depending on the requirements of the final application hosting the algorithm [38].

A cost function can be defined as the instantaneous squared error, depending on both parameters to estimate. In the complex case, the cost function to minimize is

$$J(\mathbf{w}_n, \mathbf{q}_n) = e[n]e^*[n]. \tag{27}$$

Firstly, the general form of the learning rule for the control points reads

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_{\mathrm{q}} \nabla_{\mathbf{q}_n} J(\mathbf{w}_n, \mathbf{q}_n), \tag{28}$$

where $\mu_{\mathrm{q}}[n]$ is the weight or learning rate of the update. Invoking elementary complex differentiation rules [39], while considering the other parameter $\mathbf{w}_n$ constant, the partial derivative can be written as

$$\begin{aligned}\frac{\partial J(\mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{q}_n^*} &= e[n]\frac{\partial e^*[n]}{\partial \mathbf{q}_n^*} + e^*[n]\frac{\partial e[n]}{\partial \mathbf{q}_n^*} \\ &= -e[n]\frac{\partial s^*[n]\boldsymbol{\Psi}_n^T(\mathbf{1}+\mathbf{q}_n^*)}{\partial \mathbf{q}_n^*} + 0 \\ &= -e[n]s^*[n]\boldsymbol{\Psi}_n. \end{aligned} \tag{29}$$

Substituting in (28), the final learning rule reads

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_{\mathrm{q}} e[n] s^*[n] \boldsymbol{\Psi}_n. \tag{30}$$

Secondly, the learning rule for the filter $\mathbf{w}_n$ is

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{\mathrm{w}} \nabla_{\mathbf{w}_n} J(\mathbf{w}_n, \mathbf{q}_n), \tag{31}$$

where $\mu_{\mathrm{w}}[n]$ is the learning rate of the update. The partial derivative with respect to $\mathbf{w}_n$, holding $\mathbf{q}_n$ constant, is given as

$$\begin{aligned}\frac{\partial J(\mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{w}_n^*} &= e[n]\frac{\partial e^*[n]}{\partial \mathbf{w}_n^*} + e^*[n]\frac{\partial e[n]}{\partial \mathbf{w}_n^*} \\ &= -e[n]\frac{\partial s^*[n]\boldsymbol{\Psi}_n^T(\mathbf{1}+\mathbf{q}_n^*)}{\partial \mathbf{w}_n^*} + 0 \\ &= -e[n]\boldsymbol{\Psi}_n^T(\mathbf{1}+\mathbf{q}_n^*)\mathbf{x}_n^* \\ &\quad - \frac{s[n]\mathbf{x}_n^*}{|s[n]|}\mathrm{Re}\left\{e[n]\dot{y}[n]\right\}. \end{aligned} \tag{32}$$

Here, $\dot{y}[n] = s^*[n]\dot{\mathbf{u}}_n^T \mathbf{C}_P(\mathbf{1}+\mathbf{q}_{i,n}^*)$, and the vector $\dot{\mathbf{u}}_n \in \mathbb{R}^{(P+1)\times 1} = [P\mathbf{u}_n^{P-1} \ (P-1)\mathbf{u}_n^{P-2} \ \cdots \ 1 \ 0]^T$ represents the derivative of the abscissa vector $\mathbf{u}_n$. Substituting this result in (31), the final learning rule reads

$$\begin{aligned}\mathbf{w}_{n+1} &= \mathbf{w}_n + \mu_{\mathrm{w}}\mathbf{x}_n^* \\ &\quad \times \left(e[n]\boldsymbol{\Psi}_n^T(\mathbf{1}+\mathbf{q}_n^*) + \frac{\mathrm{Re}\left\{e[n]\dot{y}[n]s[n]\right\}}{|s[n]|}\right). \end{aligned} \tag{33}$$

B. Complexity of the Proposed SPW Model

As described above, the algorithm is designed to be executed sample by sample, and thus this analysis is presented in terms of FLOPs per sample. This metric basically collects the numbers of multiplications, additions and subtractions to be executed. Furthermore, the number of real multiplications required by the algorithm is also presented separately, as it can constitute an important metric for hardware implementations, such as FPGAs, where resources are limited.

The computational complexity of this method is detailed based on the model identification and coefficient update stages done in each iteration, following the steps below:

1) Filter the input signal to get $s[n]$
2) Compute index and abscissa parameters i_n, u_n, and $\mathbf{u}_n$
3) Perform spline interpolation to obtain $y[n]$
4) Update coefficient $\mathbf{q}_{n+1}$
5) Update coefficient $\mathbf{w}_{n+1}$

The corresponding complexity expressions are collected in Table II , giving the upper bounds for the final complexity, as in a hardware implementation many trivial operations such as multiplications with zero elements or integer powers of 2 or 1/2 do not reflect any added complexity. The complexity expressions are presented as a function of the model parameters P, M_{w} and Q.

V. Complex Spline-Based Adaptive Wiener-Hammerstein Model

The Wiener-Hammerstein model is a combination of the Wiener and Hammerstein models, capable of modeling more complex nonlinear systems with memory. This approach cascades a second linear filter to the Wiener structure to compensate for memory effects appearing after the nonlinearity [30]. Again, the nonlinear function is built upon B-spline interpolation, and it is injected to the linear intermediate signal $l[n]$, as indicated in Fig. 5. We refer to this approach as the SPWH model.

According to the nomenclature in Fig. 5, the intermediate signals, $l[n]$ and $s[n]$, and the model output signal, $y[n]$, can be defined as in (24) and (16). Hence,

$$l[n] = \mathbf{w}_n^T \mathbf{x}_n, \tag{34}$$

$$s[n] = l[n] + l[n]\boldsymbol{\Psi}_n^T \mathbf{q}_n, \tag{35}$$

$$y[n] = \mathbf{g}_n^T \mathbf{s}_n, \tag{36}$$

where the new filters are defined now as $\mathbf{w}_n \in \mathbb{C}^{M_{\mathrm{w}}\times 1} = [w[0] \quad w[1] \ \cdots \ w[M_{\mathrm{w}}-1]]^T$, $\mathbf{g}_n \in \mathbb{C}^{M_{\mathrm{g}}\times 1} = [g[0] \ g[1] \ \cdots \ g[M_{\mathrm{g}}-1]]^T$, and $\mathbf{x}_n$ and $\mathbf{s}_n$ are the regression signals for $x[n]$ and $s[n]$, respectively.

A. Learning Rules

In order to obtain the learning rules for the SPWH model, the same gradient descent approach can be adopted, where an error signal is used to generate a cost function which is minimized with respect to the coefficients to be estimated. In this case, three different coefficient vectors need to be updated, namely the spline control points $\mathbf{q}_n$, and the two linear filters, $\mathbf{w}_n$, and $\mathbf{g}_n$.

The cost function for the SPWH adaptation can then be written as

$$J(\mathbf{w}_n, \mathbf{q}_n, \mathbf{g}_n) = e[n]e^*[n], \tag{39}$$

and the learning expressions are then

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{\mathrm{w}} \nabla_{\mathbf{w}_n} J(\mathbf{w}_n, \mathbf{q}_n, \mathbf{g}_n), \tag{40}$$

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_{\mathrm{q}} \nabla_{\mathbf{q}_n} J(\mathbf{w}_n, \mathbf{q}_n, \mathbf{g}_n), \tag{41}$$

$$\mathbf{g}_{n+1} = \mathbf{g}_n - \mu_{\mathrm{g}} \nabla_{\mathbf{g}_n} J(\mathbf{w}_n, \mathbf{q}_n, \mathbf{g}_n), \tag{42}$$

where μ_{w}, μ_{q}, and μ_{g} represent the learning rates for each update. The complex gradient approach is again used to calculate the derivatives, assuming that the other coefficients are

TABLE II
NUMBER OF REQUIRED ARITHMETIC OPERATIONS IN EACH ITERATION OF THE SPW IDENTIFICATION ALGORITHM, IN TERMS OF FLOPS PER SAMPLE, AND REAL MULTIPLICATIONS PER SAMPLE

Operation		FLOPs/sample	Real multiplications/sample
Model identification	$s[n]$	$8M_{\mathrm{w}}-2$	$4M_{\mathrm{w}}$
	$y[n]$	$2P^2+7P+18$	$P^2+4P+8+sqrt$
	Total	$2P^2+7P+8M_{\mathrm{w}}+16$	$P^2+4P+4M_{\mathrm{w}}+8+sqrt$
Coefficient update	$\mathbf{w}_{n+1}$	$2P^2+8P+10M_{\mathrm{w}}+17+div$	$P^2+3P+6M_{\mathrm{w}}+15+div$
	$\mathbf{q}_{n+1}$	$2P+2Q+10$	$2P+8$
	Total	$2P^2+10P+2Q+10M_{\mathrm{w}}+27+div$	$P^2+5P+6M_{\mathrm{w}}+23+div$
Total		$4P^2+17P+2Q+18M_{\mathrm{w}}+43+div$	$2P^2+9P+10M_{\mathrm{w}}+31+div+sqrt$

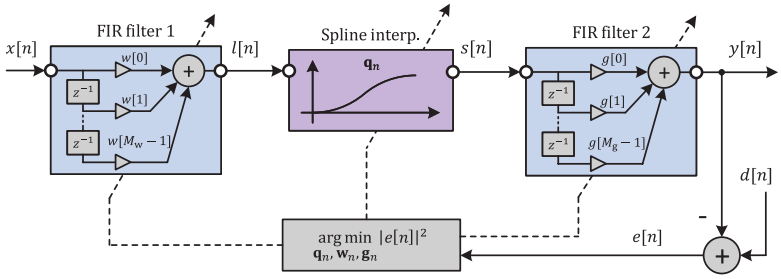


Fig. 5. Architecture of the complex spline interpolated Wiener-Hammerstein system. Note that the *spline interpolation* block incorporates the scheme presented in Fig. 1.

constants. Thus, the first gradient in (40) can be obtained as

$$\begin{aligned}\frac{\partial J(\mathbf{w}_n,\mathbf{q}_n,\mathbf{g}_n)}{\partial \mathbf{w}_n^*} &= e[n]\frac{\partial e^*[n]}{\partial \mathbf{w}_n^*}+e^*[n]\frac{\partial e[n]}{\partial \mathbf{w}_n^*}\\ &= -e[n]\frac{\partial \mathbf{g}_n^H\mathbf{s}_n^*}{\partial \mathbf{w}_n^*}-e^*[n]\frac{\partial \mathbf{g}_n^T\mathbf{s}_n}{\partial \mathbf{w}_n^*}\\ &= \cdots\\ &= e[n]\boldsymbol{\Sigma}_n\mathbf{g}_n^*+\boldsymbol{\Upsilon}_n\mathrm{Re}\left\{e[n]\dot{\mathbf{Y}}_n\mathbf{g}_n^*\right\}, \quad (43)\end{aligned}$$

where $\dot{\mathbf{Y}}_n=\mathrm{diag}\{\dot{y}[n],\ \ldots,\ \dot{y}[n-M_{\mathrm{g}}+1]\}$, and $\boldsymbol{\Sigma}_n$ and $\boldsymbol{\Upsilon}_n$ are presented in (37) and (38), as shown at the bottom of this page, respectively.

The final learning update for $\mathbf{w}_n$ can then be written as

$$\mathbf{w}_{n+1}=\mathbf{w}_n+\mu_{\mathrm{w}}\left(e[n]\boldsymbol{\Sigma}_n\mathbf{g}_n^*+\boldsymbol{\Upsilon}_n\mathrm{Re}\left\{e[n]\dot{\mathbf{Y}}_n\mathbf{g}_n^*\right\}\right), \quad (44)$$

Note that removing the filter $\mathbf{g}_n$ ($\mathbf{g}_n=[1\ 0\ \cdots\ 0]^T$) leads to the learning rule of the SPW model, presented in (44). Additionally, further removing the nonlinearity ($\mathbf{q}_n=[0\ 0\ \cdots\ 0]^T$) leads to the expression of the classical LMS filter, shown in (19).

When considering $\mathbf{g}_n$ and $\mathbf{q}_n$, they directly correspond to the learning rules obtained in the SPH case, due to the SPWH structure. Hence, they can directly be written as

$$\mathbf{g}_{n+1}=\mathbf{g}_n+\mu_{\mathrm{g}}e[n]\mathbf{s}_n^*, \quad (45)$$

$$\mathbf{q}_{n+1}=\mathbf{q}_n+\mu_{\mathrm{q}}e[n]\,\boldsymbol{\Xi}_n^T\mathbf{L}_n^*\mathbf{g}_n^*, \quad (46)$$

where $\mathbf{L}_n$ is a diagonal matrix containing the regression of $l[n]$. Note that in the learning rule for $\mathbf{q}_{n+1}$, it is again assumed that the rate of change of $\mathbf{q}_n$ over the span of the filter length M_{g} is negligible, i.e., $\mathbf{q}_n\approx\mathbf{q}_{n+M_{\mathrm{g}}}$.

B. Complexity of the Proposed SPWH Model

The complexity of the proposed SPWH model is analyzed in terms of FLOPs and real multiplications per sample.

It is again important to remark the relatively high complexity involved in calculating the term $\boldsymbol{\Xi}_n^T\mathbf{L}_n^*\mathbf{w}_n^*$ in the update of $\mathbf{q}_{n+1}$, as $\boldsymbol{\Xi}_n$ is a $Q\times M_{\mathrm{w}}$ matrix, and $\mathbf{w}_n$ may contain a large number of taps. To ease this update, we propose a similar complexity reduction approximation as was done for the SPH model, where only a specific temporal span of the matrix $\boldsymbol{\Xi}_n$ (row dimension) is chosen for the update. Thus, only τ taps, i,e,. rows of $\boldsymbol{\Xi}_n$, $\mathbf{w}_n$, and $\mathbf{L}_n$ can be selected for the learning rule, simplifying the overall update while having minimal effect on the performance. The approximation is similar to the SPH model, and is given in (21).

$$\boldsymbol{\Sigma}_n=\left[\boldsymbol{\Psi}_n(\mathbf{1}+\mathbf{q}_n^*)\mathbf{x}_n^*\quad \boldsymbol{\Psi}_{n-1}(\mathbf{1}+\mathbf{q}_{n-1}^*)\mathbf{x}_{n-1}^*\quad \cdots\quad \boldsymbol{\Psi}_{n-M_{\mathrm{g}}+1}(\mathbf{1}+\mathbf{q}_{n-M_{\mathrm{g}}+1}^*)\mathbf{x}_{n-M_{\mathrm{g}}+1}^*\right], \quad (37)$$

$$\boldsymbol{\Upsilon}_n=\left[\frac{l[n]\,\mathbf{x}_n^*}{|l[n]|}\quad \frac{l[n-1]\,\mathbf{x}_{n-1}^*}{|l[n-1]|}\quad \cdots\quad \frac{l[n-M_{\mathrm{g}}+1]\,\mathbf{x}_{n-M_{\mathrm{g}}+1}^*}{|l[n-M_{\mathrm{g}}+1]|}\right]. \quad (38)$$

TABLE III
NUMBER OF REQUIRED ARITHMETIC OPERATIONS IN EACH ITERATION OF THE SPWH IDENTIFICATION ALGORITHM, IN TERMS OF FLOPS PER SAMPLE, AND REAL MULTIPLICATIONS PER SAMPLE

Operation		FLOPs/sample	Real multiplications/sample
Model identification	$l[n]$	$8M_{\mathrm{w}} - 2$	$4M_{\mathrm{w}}$
	$s[n]$	$2P^2 + 7P + 19$	$P^2 + 4P + 8 + sqrt$
	$y[n]$	$8M_{\mathrm{g}} - 2$	$4M_{\mathrm{g}}$
	Total	$2P^2 + 7P + 8(M_{\mathrm{w}} + M_{\mathrm{g}}) + 15$	$P^2 + 4P + 4(M_{\mathrm{w}} + M_{\mathrm{g}}) + 8 + sqrt$
Coefficient updates	$\mathbf{w}_{n+1}$	$2P^2 + 6P + 12M_{\mathrm{w}}M_{\mathrm{g}} + 20M_{\mathrm{w}} + 14M_{\mathrm{g}} + 7 + div$	$P^2 + 3P + 6M_{\mathrm{w}}M_{\mathrm{g}} + 14M_{\mathrm{w}} + 8M_{\mathrm{g}} + 9 + div$
	$\mathbf{q}_{n+1}$	$4\tau Q + 6\tau + 6Q + 2$	$4P + 2P\tau + 6\tau + 6$
	$\mathbf{g}_{n+1}$	$8M_{\mathrm{g}} + 2$	$4M_{\mathrm{g}} + 2$
	Total	$2P^2 + 6P + 4\tau Q + 6\tau + 6Q$ $+12M_{\mathrm{w}}M_{\mathrm{g}} + 20M_{\mathrm{w}} + 22M_{\mathrm{g}} + 11$	$P^2 + 7P + 2P\tau + 6\tau + 6M_{\mathrm{w}}M_{\mathrm{g}}$ $+14M_{\mathrm{w}} + 12M_{\mathrm{g}} + 17 + div$
Total	-	$4P^2 + 13P + 4\tau Q + 6\tau + 6Q$ $+12M_{\mathrm{w}}M_{\mathrm{g}} + 28M_{\mathrm{w}} + 30M_{\mathrm{g}} + 26 + div$	$2P^2 + 11P + 2P\tau + 6\tau + 6M_{\mathrm{w}}M_{\mathrm{g}}$ $+18M_{\mathrm{w}} + 16M_{\mathrm{g}} + 25 + sqrt + div$

The upper bound complexity expressions for the SPWH model are gathered in Table III. Note that the case $\tau = M_{\mathrm{w}}$ corresponds to the original learning update, with no complexity reduction method applied.

VI. APPLICATIONS, USE CASES, AND RF MEASUREMENTS

The proposed modeling approaches can be applied to many use cases and applications. In this section, we specifically focus on DSIC and DPD as direct and inverse modeling problems, respectively. We present extensive RF measurement results to verify and demonstrate the performance of the proposed techniques.

In the results, the widely used MP and GMP models are adopted for reference, as they are the most common high-performance techniques used in the literature [9], [16], [23], [40]. The input/output relation of the GMP model can be expressed as

$$y[n] = \sum_{\substack{p=1 \\ p\ \mathrm{odd}}}^{P_{\mathrm{MP}}} \sum_{m=0}^{M_{\mathrm{MP}}} \alpha_{p,m} x[n-m]|x[n-m]|^{p-1} + \sum_{\substack{p=3 \\ p\ \mathrm{odd}}}^{P_{\mathrm{C}}} \sum_{m=0}^{M_{\mathrm{C}}} \sum_{\substack{l=-L_{\mathrm{C}} \\ l\neq 0}}^{L_{\mathrm{C}}} \beta_{p,m,l} x[n-m]|x[n-m+l]|^{(p-1)}, \tag{47}$$

where P_{MP}, M_{MP}, and $\alpha_{p,m}$ represent the nonlinear order, memory, and model coefficients of the aligned envelope terms. Additionally, P_{C}, M_{C}, L_{C}, and $\beta_{p,m,l}$ are the nonlinear order, memory, depth, and model coefficients of the lagging and leading terms. For simplicity, we assume in this work the same number of lagging and leading coefficients, denoted here by $C = \lfloor P_{\mathrm{C}}/2 \rfloor M_{\mathrm{C}} L_{\mathrm{C}}$. In fact, C indicates the number of additional coefficients compared to the MP model. Obviously, the MP model can be directly obtained from (47) by setting the lagging and leading coefficients to zero (and thus $C = 0$). In both models, the set of coefficients can be learned using the LMS algorithm, thus facilitating a fair comparison with the proposed algorithms. The basis functions of the models, however, need to be prewhitened or orthogonalized before the processing [41], if LMS is applied. The reason for this is the poor convergence performance of the LMS algorithm if the elements of the input vector are highly correlated, stemming from the large eigenvalue spread of the input signal covariance matrix. In this case, the static basis functions can certainly be expected to be correlated, since they all depend on the original transmit signal. Orthogonalizing them is thus a necessary step to ensure efficient parameter learning by the LMS algorithm. The computational complexity of this process is included within the parameter update stage. Further details and analysis of the orthogonalized polynomial models and LMS learning in the context of DSIC can be found in [25].

A. Direct Modeling: DSIC in IBFD Transceivers

Self-interference cancellation technology has gathered a lot of attention in the past 10 years as a key technology for realizing IBFD communications [3], [23], [40]. IBFD devices are capable of transmitting and receiving simultaneously on the same channel, thus enabling a twofold increase in the spectral efficiency and data rate, without requiring any additional bandwidth. Fig. 6 shows a typical IBFD transceiver structure. Besides IBFD, DSIC technology can benefit also other communication technologies [4], as well as joint sensing and communications [42].

To make IBFD commercially viable, the problem of SI must be dealt with. SI refers to the unwanted own transmit signal that is leaked into the receiver chain, potentially saturating the receiver and masking the desired received signal. Thus, any full-duplex transceiver must be capable of removing the SI signal in an efficient manner to provide reliable communication. In practice, this is done in three stages: using propagation domain isolation techniques, analog SI cancellation, and digital cancellation (DSIC), as illustrated in Fig. 6. DSIC, which is our focus here, is a forward modeling problem, typically aiming at modeling the nonlinear transmitter and the SI propagation channel [5]. The proposed cascaded models are well suited for modeling such a system.

We demonstrate the feasibility of the proposed techniques in the context of DSIC in IBFD by utilizing measured data

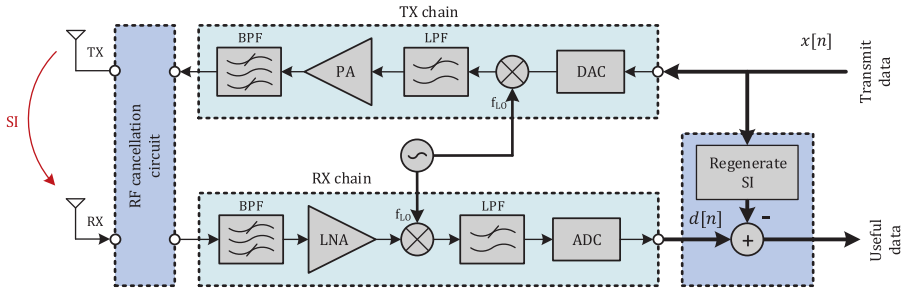


Fig. 6. Generic IBFD transceiver architecture.

TABLE IV
THE RF MEASUREMENT AND MODEL PARAMETERS IN THE DSIC EXPERIMENT. THE COMPUTATIONAL COMPLEXITY IN THE CANCELLATION AND PARAMETER UPDATE STAGES IS PRESENTED IN THE LAST TWO COLUMNS, IN TERMS OF REAL MULTIPLICATIONS PER SAMPLE. THE MP AND GMP BF ORTHOGONALIZATION COMPLEXITY IS INCLUDED IN THE COEFFICIENT UPDATE STAGE

System parameters	Value	Model	P	C	Q	$M_{\mathrm{w,pre}}/M_{\mathrm{w,post}}$	$M_{\mathrm{g,pre}}/M_{\mathrm{g,post}}$	Mul. - canc.	Mul. - update
Center frequency	2.56 GHz	SPH DSIC	3	-	5	-	25 / 50	329	380
Transmit waveform	20/40/80 MHz OFDM	SPW DSIC	3	-	5	25 / 50	-	329	497
Meas. sample size	1 Msamples	SPWH DSIC	3	-	5	0 / 10	25 / 50	369	5647
Transmit power	$\approx$ 24 dBm	MP DSIC	11	0	-	-	25 / 50	1840	833 570
RX losses	4 dB	GMP DSIC	11	12	-	-	25 / 50	1960	923 522

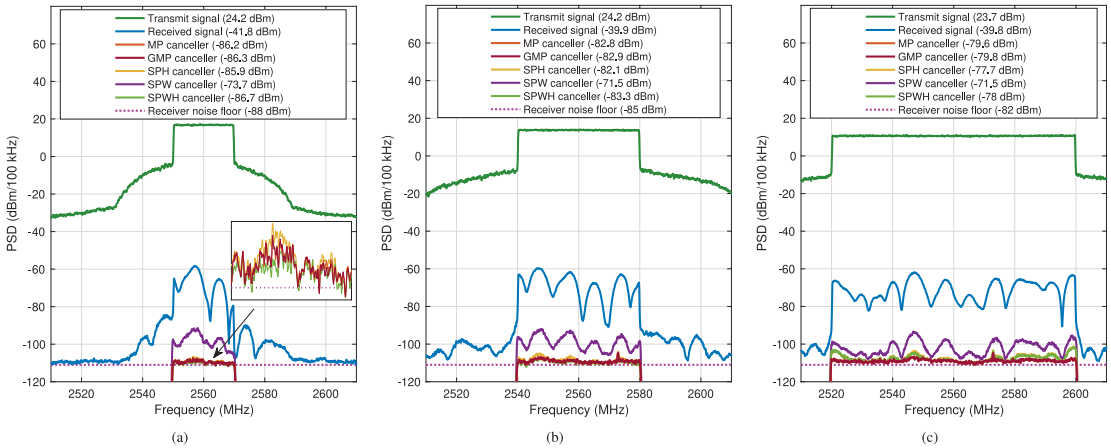


Fig. 7. PSDs of the overall signal after the different DSICs, for (a) 20 MHz, (b) 40 MHz, and (c) 80 MHz instantaneous bandwidths.

originally used in [43], where a back-to-back relay antenna was used to provide the analog isolation between the transmitter and receiver nodes, and no active analog cancellation was used.

The parameterizations for the different DSIC models are collected in Table IV, along with the computational complexity in terms of real multiplications per input sample in the cancellation and coefficient update stages. In both SPH and SPWH models, the proposed complexity reduction method is considered, with $\tau = 5$. In the GMP model, $P_{\mathrm{C}} = 5$, $M_{\mathrm{C}} = 3$, and $L_{\mathrm{C}} = 2$ are considered, which leads to $C = 12$ coefficients. Additionally, a large number of memory taps is considered in each model due to the high frequency selectivity of the SI channel. First, the power spectral densities (PSDs) of the transmit, received, and signal after cancellation with the algorithms are shown in Fig. 7, for three different instantaneous signal bandwidths of 20, 40, and 80 MHz. In all cases, the SPH and SPWH models achieve a similar cancellation to that of the MP and GMP models, regardless of the substantial complexity reduction. With the narrowest 20 MHz bandwidth, the amounts of achieved digital cancellation are 44.1 dB (SPH) and 44.9 dB (SPWH). With the widest 80 MHz bandwidth, the amounts of achieved cancellation are 37.9 dB (SPH) and 38.2 dB (SPWH). These

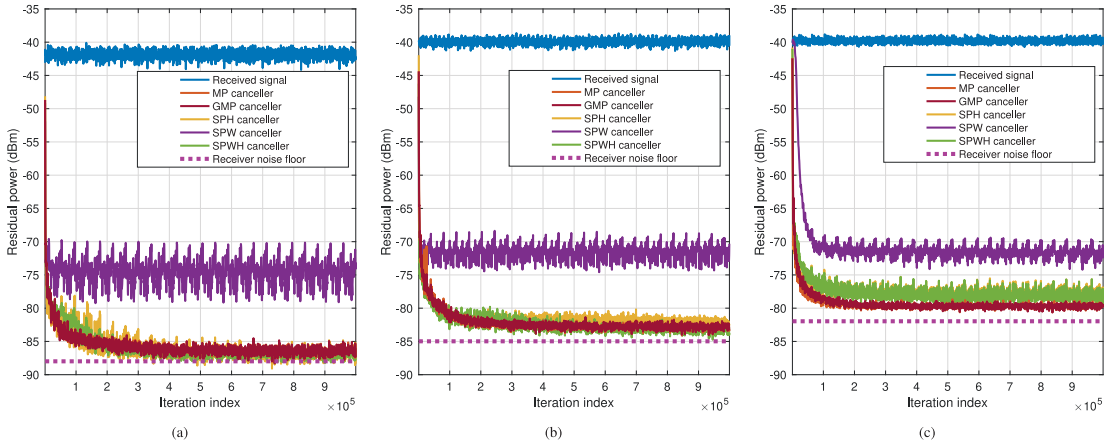


Fig. 8. Residual powers with respect to the iteration index after the different DSICs, for (a) 20 MHz, (b) 40 MHz, and (c) 80 MHz instantaneous bandwidths.

results indicate high modeling accuracy, with the SPWH model even outperforming the MP and GMP models in the narrowband cases. In this particular measurement, the SPW model yields somewhat lower performance with around 32 dB cancellation in the three considered bandwidth cases. This is expected due to the SI signal model naturally following a Hammerstein type structure. Additionally, the proposed methods are capable of reducing the cancellation complexity by more than 82% (SPH and SPW) and 80% (SPWH) in the cancellation stage with respect to the MP and GMP models. In the coefficient update stage, the computational complexity of all the methods is reduced by more than 99% compared to the MP and GMP. This is due, in large part, to them not requiring BF orthogonalization, in contrast to the polynomial-based models. These results illustrate the excellent performance-complexity trade-off of the proposed solutions.

In order to evaluate the convergence properties of the algorithms, Fig. 8 shows the residual powers of the received signals after cancellation, for the three bandwidths mentioned above. All the models are essentially achieving a similar convergence speed, despite the proposed spline based models not using orthogonalization. Altogether, the results show that excellent digital cancellation can be obtained with the proposed algorithms, regardless of the large processing complexity reduction.

B. Inverse Modeling: Digital Predistortion

In this section, the proposed models are tested in the context of DPD. DPD is a well-established technique that aims at minimizing the unwanted emissions and nonlinear distortion originated from PAs. This approach applies a nonlinear transformation to the input signal that pre-compensates for the unwanted PA effects [2]. Thus, in DPD processing, choosing a nonlinear model with the best complexity-performance trade-off for the device at hand is a key design challenge.

In this paper, we apply DPD linearization to a millimeter wave (mmW) active antenna array system operating at 28 GHz. The array linearization problem, in the single user case, is a single-input single-output identification problem, similar to regular DPD linearization of a single PA [44]–[46]. However, in a K-antenna array, there are also K PAs, each with unique (though somewhat similar) nonlinear characteristics. The DPD can thus guarantee good linearization only in the main beam direction, while in the rest of the directions, the beampattern of the array will help to keep the nonlinear distortions low [44]. Another important feature of arrays is the load modulation of the PAs, which occurs due to coupling between the antennas [47]. This will make the nonlinear characteristics of the array *beam-dependent*. Thus, any linearization solution for active arrays needs to take this into account, as the optimal DPD will also depend on the beam direction. DPD solutions which are real-time adaptive are one viable solution to track and adapt the DPD coefficients as the beam is steered. A further concern at mmWs, as the signal bandwidths are wide and matching is more problematic compared to lower frequencies, is the frequency selectivity of the transmitter. There could be, for example, distinct frequency selectivity between the baseband/IF parts and the PA, which the DPD estimator would interpret as part of the PA response, thus requiring large memory depth and unnecessary complexity in the DPD. The proposed adaptive cascaded models can *separate* these linear distortions from the nonlinear effects of the PAs, and thus offer much lower complexity linearization, which is an important criterion when designing real-time systems.

For learning, we adopt the so-called ILA, which is illustrated in Fig. 9. Here, the DPD coefficients are estimated through the post-inverse of the PA, which can be calculated directly from the PA input and output signals [9]. The post-inverse coefficients are copied to the digital predistorter, and the learning procedure is then typically iterated a few times to reach steady-state.

The DPD performance is evaluated through the well-known error vector magnitude (EVM) and normalized mean square error (NMSE) [16]. However, since an over-the-air (OTA) DPD system is considered, the out-of-band performance is measured with the total radiated power (TRP) based adjacent channel

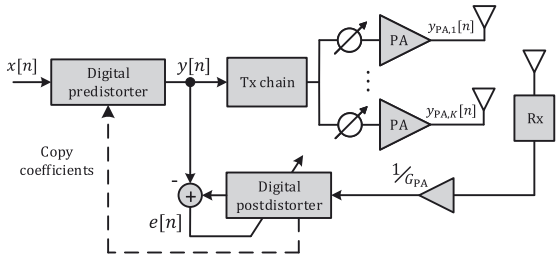


Fig. 9. Illustration of the ILA DPD system, in the context of a K-antenna array. The DPD is estimated directly from the predistorted and OTA received signals. G_{PA} represents the estimate of the complex linear gain of the Tx chain.

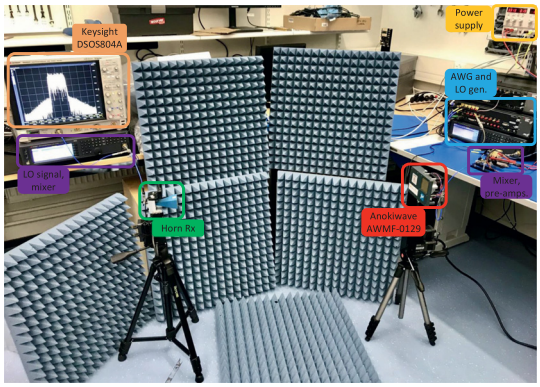


Fig. 10. The 5 G NR mmW/FR-2 RF deployment used in the DPD experiment.

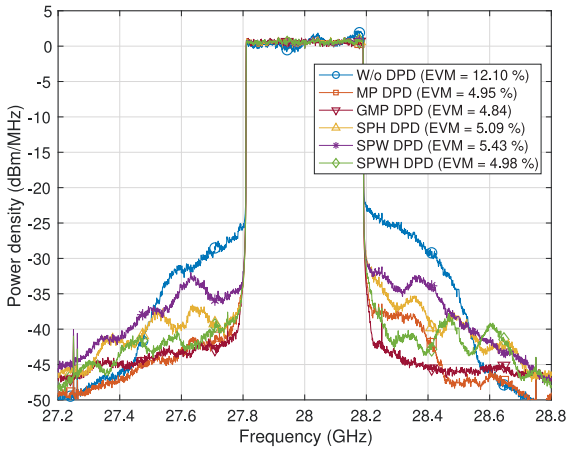


Fig. 11. Example illustration of DPD OTA linearization results at 28 GHz, with 400 MHz bandwidth and EIRP $\approx$ +42 dBm.

leakage ratio (ACLR), which is the filtered mean power centered on the assigned channel frequency to the filtered mean power centred on an adjacent channel frequency, measured by integrating the powers over the whole beamspace, while keeping the *beamforming angle* fixed [48].

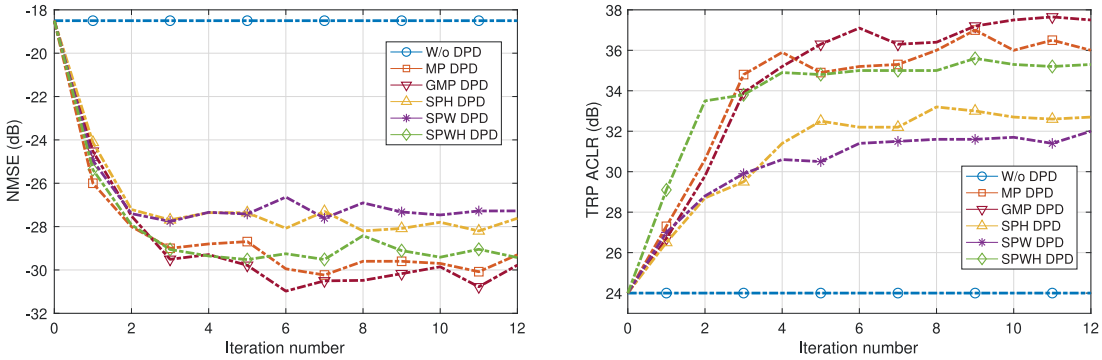
The overall array measurement setup is shown in Fig. 10, containing the test device, which is an Anokiwave AWMF-0129 64-element active antenna array, together with other instruments to facilitate signal generation and analysis at 28 GHz center frequency. Firstly, a Keysight M8190 arbitrary waveform generator (AWG) is used to generate an IF signal at 3.5 GHz. The signal is then upconverted to the 28 GHz carrier by a Keysight N5183B-MXG acting as the LO at 24.5 GHz and a Marki Microwave T31040 L mixer, and then filtered by a Marki microwave FB3300 band-pass filter (BPF). The upconverted signal is then driven through two pre-amplifiers, Analog Devices HMC499LC4 and Analog Devices HMC1131, facilitating enough power to drive the active antenna array to saturation. The array transmits the signal OTA, and the signal is captured by a horn antenna at the observation receiver, such that the receiver antenna is well aligned with the main beam. In this experiment, the *beamforming angle* is considered as 0 degrees. The signal is then carried through an attenuator and another mixer to be down-converted back to IF. Then, a Keysight DSOS804 A oscilloscope is used as the actual digitizer to facilitate the post-processing on a host PC, where the DPD algorithms are executed.

In the DPD measurements, a 5G NR FR-2 OFDM signal with 120 kHz subcarrier spacing (SCS) and 264 active resource blocks (RBs) is adopted. This configuration maps to 400 MHz channel bandwidth. The proposed DPD models are tested with 12 ILA iterations, and 50 ksamples each. The adopted model parametrization and DPD complexity are presented in Table V. The cross-terms of the GMP model are configured with $P_{\mathrm{C}} = 9$, $M_{\mathrm{C}} = 2$, $L_{\mathrm{C}} = 2$, which leads to $C = 16$ coefficients. A snap-shot linearization example is showed in Fig. 11, at effective isotropic radiated power (EIRP) of +42 dBm. Within the three proposed solutions, the SPWH model obtains the best linearization performance, and the closest to the reference MP and GMP solutions. This model is followed by the SPH DPD, which also obtains a good DPD linearization, slightly degraded but close to SPWH. The SPW model follows somewhat behind the other DPD solutions, but it is still capable of obtaining a fair amount of performance. This result can be explained as follows. On the one hand, the SPW model cannot correct memory effects and distortion appearing in the mixing and filtering stage (T31040 L MX and FB3300 BPF), before the driver PA, leading to performance loss. On the other hand, the SPH model cannot properly account for the frequency selectivity appearing in the OTA transmission. The SPWH model, in turn, is able to account for both phenomena, achieving thus an enhanced performance compared to the other models.

The same conclusion can be made from Fig. 12, which shows the NMSE and TRP ACLR performance as a function of the iteration number. It can also be observed that the convergence of the DPD models is fast in general, reaching the steady-state in approximately 4-5 DPD iterations. The 5G NR ACLR limit of 28 dBc, measured using the TRP [48] approach, is fulfilled in all cases. It can be seen from Table V that the computational complexity with respect to the classical MP is also greatly

TABLE V
THE RF MEASUREMENT AND MODEL PARAMETERS IN THE DPD EXPERIMENT. THE COMPUTATIONAL COMPLEXITY OF THE DPD MAIN PATH AND DPD LEARNING STAGES IS PRESENTED IN THE LAST TWO COLUMNS, IN TERMS OF REAL MULTIPLICATIONS PER SAMPLE. THE MP AND GMP BF ORTHOGONALIZATION COMPLEXITY IS INCLUDED IN THE DPD LEARNING STAGE

System parameters	Value	SI model	P	C	Q	$M_{\mathrm{w,pre}}/M_{\mathrm{w,post}}$	$M_{\mathrm{g,pre}}/M_{\mathrm{g,post}}$	Mul. - main path	Mul. - learning
Center frequency	28 GHz	SPH DPD	3	-	5	-	0 / 15	89	140
Transmit waveform	400 MHz NR OFDM	SPW DPD	3	-	5	0 / 15	-	89	137
SCS / RBs	120 kHz / 264	SPWH DPD	3	-	5	0 / 13	0 / 13	133	1459
PAPR (@ 0.01%)	7 dB	MP DPD	11	0	-	-	0 / 10	280	17 970
EIRP	+42 dBm	GMP DPD	11	16	-	-	0 / 10	400	33 162



(a) (b)

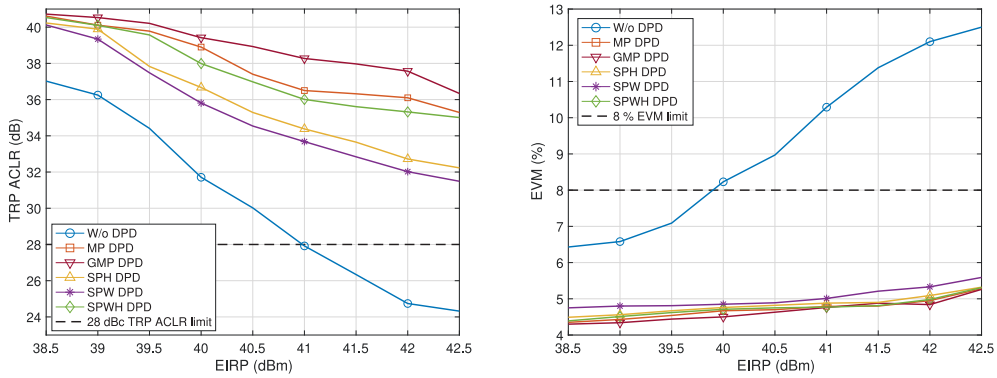
Fig. 12. DPD performance in terms of (a) NMSE, and (b) TRP ACLR of the proposed DPD models as a function of the iteration index, with 400 MHz bandwidth and EIRP $\approx$ +42 dBm.

reduced by 68% (SPH and SPW) and 53% (SPWH) in the DPD main path. With respect to GMP, the complexity is reduced by 78% (SPH and SPW) and 67% (SPWH). In the DPD learning, the computational complexity of the proposed models is reduced by more than 90% with respect to MP, and by more than 95% with respect to GMP. These results verify the excellent performance-complexity trade-off of the proposed techniques.

To further demonstrate the performance of the proposed models, a third and final experiment featuring a power sweep is presented. The power sweep shows the ACLR, measured through the TRP, and EVM performance metrics as a function of the EIRP, which ranges from +38.5 dBm to +42.5 dBm. With this experiment, we try to shed some light into two main things. First, to evaluate whether the TRP ACLR or the EVM is the limiting performance metric in the mmW system, in terms of maximum achievable EIRP. Second, to assess the performance of the proposed and reference DPD models as the radiated power of the antenna array varies. The obtained TRP ACLR and EVM results are presented in Fig. 13. Firstly, it can be clearly seen that, when no predistortion is applied, the EVM constitutes the limiting factor in terms of maximum achievable EIRP, basically surpassing the 8% EVM limit [48] at approximately +39.8 dBm, where the TRP ACLR limit is still fulfilled. However, when the proposed predistortion techniques are applied, both EVM and TRP ACLR limits are satisfied at least up until +42.5 dBm, and clearly somewhat beyond, as the figure shows. These findings indicate an increase in the overall power efficiency of the transmit chain, as the antenna array can be operated closer to saturation thanks to the DPD operation. Secondly, the power sweep also indicates that the DPD algorithms behave in a similar manner as commented before. The proposed SPWH model is capable of achieving a similar linearization performance to that of the MP and GMP models, despite the reduced processing and learning complexities. It is then followed by the SPH and SPW models, whose performance lie somewhat behind, but still fulfilling the specified 3GPP limits. In general, the proposed algorithms achieve good amounts of linearization performance, and successfully satisfy the 3GPP specifications [48] at least up to EIRP of +42.5 dBm and clearly beyond, while allowing for excellent processing and learning complexity reductions. These results further confirm the very favorable complexity-performance trade-offs of the proposed solutions in different scenarios.

VII. CONCLUSION

In this paper, a class of low-complexity cascaded models and learning algorithms were proposed for modeling of complex-valued nonlinear systems with memory. Hammerstein, Wiener, and Wiener-Hammerstein models were constructed, building on uniform spline-interpolated LUTs as the elementary nonlinear functions and FIR filters for the linear blocks, and gradient descent based adaptive algorithms were derived for each of the models. This approach was shown to yield remarkably simple



(a) (b)

Fig. 13. DPD OTA linearization performance at 28 GHz with 400 MHz bandwidth, as a function of the EIRP of the proposed DPD models in terms of (a) TRP ACLR, and (b) EVM.

yet effective solutions, suitable for real-time implementation and continuous learning. In order to validate and verify the proposed methods, RF measurements were carried out in the context of full-duplex self-interference cancellation and digital predistortion. First, the received signal from a complete full-duplex prototype environment was used to test the SI cancellation capability of the proposed methods, achieving high cancellation levels with much lower complexity compared to state-of-the-art. Second, a millimeter-wave active antenna array operating at 28 GHz was linearized with the proposed techniques, reducing the unwanted emissions to levels below the specified limits. These results illustrate the excellent complexity-performance trade-off of the proposed methods, as they can offer modeling performance very close to state-of-the-art regardless of the substantial complexity reduction.

REFERENCES

[1] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*, 2nd ed. Norwood, MA, USA: Artech House, 2006.

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

[3] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[4] S. Hong *et al.*, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.

[5] D. Korpi, "Full-duplex wireless: Self-interference modeling, digital cancellation, and system studies," Ph.D. dissertation, Lab. Electron. Commun. Eng., Tampere Univ. Technol., Dec. 2017.

[6] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Sel. Areas Commun.*, vol. SAC-1, no. 1, pp. 57–62, Jan. 1983.

[7] M. Ibnkahla, N. Bershad, J. Sombrin, and F. Castanie, "Neural network modeling and identification of nonlinear channels with memory: Algorithms, applications, and analytic models," *IEEE Trans. Signal Process.*, vol. 46, no. 5, pp. 1208–1220, 1998.

[8] A. Zhu, J. C. Pedro, and T. C. 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.

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

[10] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Trans. Microw. Theory Techn.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.

[11] P. Pascual Campo, D. Korpi, L. Anttila, and M. Valkama, "Nonlinear digital cancellation in full-duplex devices using spline-based hammerstein model," in *Proc. IEEE Global Commun. Conf.*, Dec. 2018.

[12] L. Ding, R. Raich, and G. Tong Zhou, "A hammerstein predistortion linearization design based on the indirect learning architecture," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process.*, vol. 3, May 2002, pp. 2689–2692.

[13] X. Hong, Yu Gong, and Sheng Chen, "A wiener model for memory high power amplifiers using b-spline function approximation," in *Proc. 17th Int. Conf. Digit. Signal Process.*, Jul. 2011, pp. 1–5.

[14] X. Wu, N. Zheng, X. Yang, J. Shi, and H. Chen, "A spline-based hammerstein predistortion for 3 G power amplifiers with hard nonlinearities," in *Proc. 2nd Int. Conf. Future Comput. Commun.*, vol. 3, May 2010, pp. 741–745.

[15] S. Chen, X. Hong, J. Gao, and C. J. Harris, "Complex-valued b-spline neural networks for modeling and inverting hammerstein systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 9, pp. 1673–1685, Sep. 2014.

[16] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.

[17] X. Hong and S. Chen, "Modeling of complex-valued wiener systems using b-spline neural network," *IEEE Trans. Neural Netw.*, vol. 22, no. 5, pp. 818–825, May 2011.

[18] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Augmented hammerstein predistorter for linearization of broad-band wireless transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 4, pp. 1340–1349, Jun. 2006.

[19] H. Kang, Y. Cho, and D. Youn, "On compensating nonlinear distortions of an OFDM system using an efficient adaptive predistorter," *IEEE Trans. Commun.*, vol. 47, no. 4, pp. 522–526, Apr. 1999.

[20] T. Wang and J. Ilow, "Compensation of nonlinear distortions with memory effects in OFDM transmitters," in *Proc. IEEE Global Telecommun. Conf.*, 2004, pp. 2398–2403.

[21] A. Sano and L. Sun, "Identification of hammerstein-wiener system with application to compensation for nonlinear distortion," in *Proc. 41st SICE Annu. Conf.*, 2002, pp. 1521–1526.

[22] M. Cheong, S. Werner, M. Bruno, J. Figueroa, J. Cousseau, and R. Wichman, "Adaptive piecewise linear predistorters for nonlinear power amplifiers with memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 7, pp. 1519–1532, Jul. 2012.

[23] D. Bharadia, E. McMilin, and S. Katti, "Full Duplex Radios," in *Proc. ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, 2013, pp. 375–386.

[24] L. Anttila, D. Korpi, V. Syrjälä, and M. Valkama, "Cancellation of power amplifier induced nonlinear self-interference in full-duplex transceivers," in *Proc. 47th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2013, pp. 1193–1198.

[25] D. Korpi, Y.-S. Choi, T. Huusari, S. Anttila, L. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile in-band full-duplex radio: Algorithms and RF measurements," in *Proc. IEEE Global Commun. Conf.*, Dec. 2015, pp. 1–7.
[26] K. Komatsu, Y. Miyaji, and H. Uehara, "Frequency-domain hammerstein self-interference canceller for in-band full-duplex OFDM systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2017, pp. 1–6.
[27] P. P. Campo *et al.*, "Gradient-adaptive spline-interpolated lut methods for low-complexity digital predistortion," *IEEE Trans. Circuits Syst. I, Reg. Papers*, to be published, doi: 10.1109/TCSI.2020.3034825.
[28] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Nonlinear spline adaptive filtering," *Signal Process.*, vol. 93, no. 4, pp. 772–783, 2013.
[29] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties," *Signal Process.*, vol. 100, pp. 112–123, 2014.
[30] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Novel cascade spline architectures for the identification of nonlinear systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 7, pp. 1825–1835, Jul. 2015.
[31] C. De Boor, *Practical Guide to Splines*. New York, NY, USA: Springer-Verlag, 1978.
[32] M. Vetterli, J. Kovaçević, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
[33] H. Prautzsch, W. Boehm, and M. Paluszny, *Bèzier and B-Spline Techniques*. Berlin, Germany: Springer, 2013.
[34] N. Safari, N. Holte, and T. Roste, "Digital predistortion of power amplifiers based on spline approximations of the amplifier characteristics," in *Proc. IEEE 66th Veh. Technol. Conf.*, Sep. 2007, pp. 2075–2080.
[35] J. K. Cavers, "Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements," *IEEE Trans. Veh. Technol.*, vol. 39, no. 4, pp. 374–382, Nov. 1990.
[36] M. O'Droma, S. Meza, and Y. Lei, "New modified saleh models for memoryless nonlinear power amplifier behavioural modelling," *IEEE Commun. Lett.*, vol. 13, no. 6, pp. 399–401, Jun. 2009.
[37] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Hoboken, NJ, USA: Wiley, 2005, vol. 40, pp. 43–46.
[38] P. S. R. Diniz *et al.*, "Adaptive Filtering: Algorithms and Practical Implementation," *The International Series in Engineering and Computer Science*, pp. 147–150, 2008.
[39] D. P. Mandic and V. Su LeeGoh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Hoboken, NJ, USA: Wiley, 2009.
[40] D. Korpi *et al.*, "Full-duplex mobile device: Pushing the limits," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 80–87, Sep. 2016.
[41] D. Korpi, Y. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile in-band full-duplex radio: Algorithms and RF measurements," in *Proc. 2015 Global Commun. Conf.*, 2015, pp. 1–7.
[42] C. B. Barneto *et al.*, "Full-duplex OFDM radar with LTE and 5 G NR waveforms: Challenges, solutions, and measurements," *IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 10, pp. 4042–4054, Oct. 2019.
[43] D. Korpi, M. Heino, C. Icheln, K. Haneda, and M. Valkama, "Compact inband full-duplex relays with beyond 100 dB self-interference suppression: Enabling techniques and field measurements," *IEEE Trans. Antennas Propag.*, vol. 65, pp. 960–965, Feb. 2017.
[44] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital predistortion for hybrid MIMO transmitters," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445–454, Jun. 2018.
[45] A. Brihuega *et al.*, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, Sep. 2020.
[46] P. P. Campo, V. Lampu, A. Meirhaeghe, J. Boutellier, L. Anttila, and M. Valkama, "Digital predistortion for 5G small cell: GPU implementation and RF measurements," *J. Signal Process. Syst.*, pp. 1–12, 2019.
[47] 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.
[48] 3GPP Tech. Spec. 38.141-2, NR; Base Station (BS) Conformance Testing, Part 2, v15.1.0 (*Release 15*), Mar. 2019.

Pablo Pascual Campo (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications and electrical engineering in 2016 and 2018, respectively, from Universidad Politécnica de Madrid, Madrid, Spain. He is currently working toward the D.Sc. degree with Tampere University, Department of Electrical Engineering, Tampere, Finland. His research interests include digital predistortion, full-duplex systems and applications, and signal processing for wireless communications at the mmW bands.

Lauri Anttila (Member, IEEE) received the D.Sc. (Tech.) degree (with distinction) in 2011 from Tampere University of Technology (TUT), Finland. Since 2016, he has been a University Researcher with the Department of Electrical Engineering, Tampere University (formerly TUT). In 2016–2017, he was a Visiting Research Fellow with Aalto University, Finland. His research interests include radio communications and signal processing, with a focus on the radio implementation challenges in systems such as 5G, full-duplex radio, and large-scale antenna systems.

Dani Korpi received the D.Sc. degree (Hons.) in electrical engineering from the Tampere University of Technology (TUT), Finland, in 2017. He is currently a Senior Specialist with Nokia Bell Labs in Espoo, Finland. He received the Best Dissertation Award of the year in TUT, as well as the Finnish Technical Sector's Award for the Best Doctoral Dissertation of 2017. His research interests include inband full-duplex radios, machine learning for wireless communications, and beyond 5G radio systems.

Mikko Valkama (Senior Member, IEEE) received the D.Sc. (Tech.) degree (with honors) from the Tampere University of Technology (TUT), Finland, in 2001. In 2003, he was a Visiting Postdoc Research Fellow with SDSU, San Diego, CA. He is currently a Full Professor and Department Head of Electrical Engineering with the newly formed Tampere University (TAU), Finland. His research interests include radio communications, radio localization, and radio-based sensing, with emphasis on 5G and beyond mobile radio networks.

PUBLICATION 4

Inverse covariance matrix estimation for low-complexity closed-loop DPD systems: methods and performance

P. Pascual Campo, L. Anttila, V. Lampu, Y. Guo, N. Wang and M. Valkama

IEEE Transactions on Microwave Theory and Techniques, accepted for publication, 2021.(2021)

Publication reprinted with the permission of the copyright holders

Inverse Covariance Matrix Estimation for Low-Complexity Closed-Loop DPD Systems: Methods and Performance

Pablo Pascual Campo, *Graduate Student Member, IEEE*, Lauri Anttila, Vesa Lampu, Markus Allén, Yan Guo, Neng Wang, and Mikko Valkama, *Senior Member, IEEE*

***Abstract*— In this article, we study closed-loop digital predistortion (DPD) systems and associated learning algorithms. Specifically, we propose various low-complexity approaches to estimate and manipulate the inverse of the input data covariance matrix (CM) and combine them with the so-called self-orthogonalized (SO) learning rule. The inherent simplicity of the SO algorithm, combined with the proposed solutions, allows for remarkably reduced complexity in the DPD system while maintaining similar linearization performance compared to other state-of-the-art methods. This is demonstrated with thorough over-the-air (OTA) mmW measurement results at 28 GHz, incorporating a state-of-the-art 64-element active antenna array, and very wide channel bandwidths up to 800 MHz. In addition, complexity analyses are carried out, which together with the measured linearization performance demonstrates favorable performance–complexity tradeoffs in linearizing mmW active array transmitters through the proposed solutions. The techniques can find application in systems where the power amplifier (PA) nonlinearities are time-varying and thus frequent or even constant updating of the DPD is required, good examples being mmW adaptive antenna arrays as well as terminal transmitters in 5G and beyond networks.**

***Index Terms*— Array transmitters, autocorrelation function, Bussgang theorem, closed-loop systems, covariance matrix (CM), digital predistortion, Gauss–Newton (GN), parameter learning, power amplifier (PA), self-orthogonalization.**

I. Introduction

CONTEMPORARY radio communication systems, such as the recently introduced 5G new radio (NR) mobile networks, build on multicarrier modulation—most notably orthogonal frequency-division multiplexing (OFDM). Multicarrier modulations are known to provide spectrally efficient waveforms but also possess high peak-to-average power ratio (PAPR) [1], [2], which complicates the operation of power amplifiers (PAs) close to saturation. In order to control the quality of the transmit waveform while ensuring high power efficiency in the transmit chain, digital predistortion (DPD) is a well-known technical approach—see, e.g., [1], [3]–[5], and the references therein. DPD aims to suppress the unwanted out-of-band (OoB) emissions and in-band nonlinear distortion by applying specific nonlinear preprocessing to the digital transmit waveform. Especially in combination with efficient PAPR mitigation methods [6], DPD can significantly improve the transmitter power efficiency while maintaining at the same time low levels of unwanted nonlinear distortions.

A particularly timely DPD application is the linearization of active antenna arrays, being an essential technology for instance in the emerging 5G NR base stations and mobile devices, especially at the millimeter-wave (mmW) bands. These operating frequencies are also referred to as the so-called frequency range 2 (FR-2) [7] in the 3GPP terminology. Timely examples of articles focusing on this topic include, e.g., [8]–[13]. In such systems, the nonlinear distortion is known to be beam-dependent, stemming from the load-modulation phenomenon [8]. Fast DPD adaptation is thus required such that the nonlinear distortions can be suppressed, while the beam is steered. This issue, along with the extremely high processing rates and channel bandwidths at mmW frequencies, calls for reduced complexity DPD approaches and associated parameter learning algorithms.

Another relatively new DPD use case is mobile device linearization. In general, mobile device power efficiency is very important, which is why the PA is typically operated close to saturation. Thus, despite the more relaxed linearity requirements compared to base-station transmitters, linearization can be pursued for maximum power efficiency. Furthermore, in mobile devices, the transmitted waveforms are very dynamic, as the resource block allocation and output power can change on a per-slot basis [14]. On the other hand, the computational resources in mobile devices are much more limited than in base stations. Thus overall, both the adaptability and the complexity of the DPD system are becoming increasingly important in mobile devices, forming one main motivation for this work and the methods described in this article.

A wide range of DPD models can be found in the existing literature, with some of the most common approaches

Manuscript received May 11, 2021; revised August 10, 2021; accepted September 24, 2021. This work was supported by the Academy of Finland under Project 301820, Project 319994, Project 323461, Project 332361, and Project 338224. *(Corresponding author: Pablo Pascual Campo.)*

Pablo Pascual Campo, Lauri Anttila, Vesa Lampu, Markus Allén, and Mikko Valkama are with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland (e-mail: mikko.valkama@tuni.fi).

Yan Guo and Neng Wang are with the Wireless Terminal Algorithm Development Department, HiSilicon Technologies Company Ltd., Shanghai 201206, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMTT.2021.3124226.

Digital Object Identifier 10.1109/TMTT.2021.3124226

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

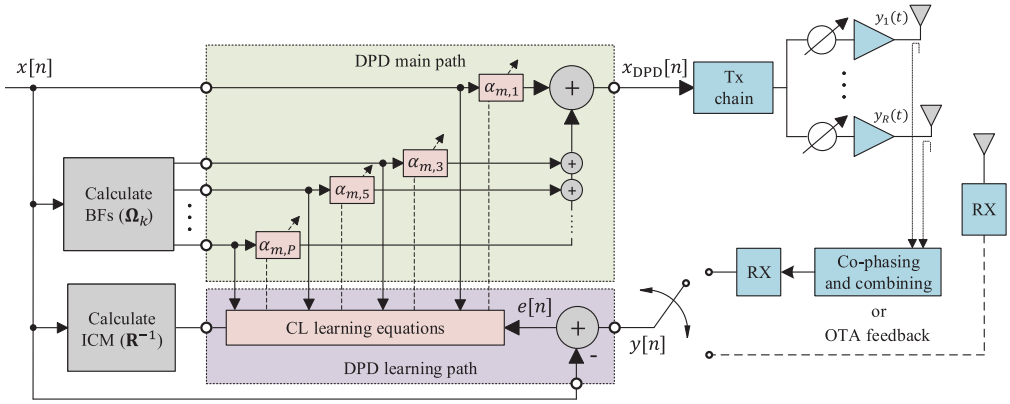


Fig. 1. Illustration of the considered closed-loop MP DPD scheme, with special emphasis on mmW active array linearization with R antenna elements and PA units. The learning signal can be obtained by either co-phasing and combining the individual PA output signals or, alternatively, by using an OTA ORX.

being the memory polynomial (MP) [1], [5], [15] and the generalized MP (GMP) [3], [15], [16]. These approaches can be seen as subsets of the Volterra series [17], [18]. In such methods, the DPD coefficients are usually estimated by using block-based regression algorithms, such as least squares (LS) or Gauss–Newton (GN) [4], [19]. These learning solutions provide reliable coefficient estimation, but they also involve relatively high computational complexity—especially if on-chip estimation were pursued. Alternatively, adaptive filtering algorithms can be used, such as least mean squares (LMS) or some of its variants [20]. The input data basis functions (BFs), however, need to be typically prewhitened or orthogonalized before the actual processing since the elements of the input data are highly correlated, leading to large eigenvalue spread of the input signal covariance matrix (CM) [21].

Literature targeting explicitly low-complexity DPD approaches and associated learning methods is, in turn, somewhat less common. Techniques aiming at this direction are, for instance, [21]–[27]. In [22], a closed-loop MP DPD model was presented, where the model coefficients were estimated with damped GN in combination with a signed regressor algorithm (SRA). However, the signed CM in the SRA suffered from rank-deficiencies, and additional Walsh-Hadamard matrix transformations were required, further increasing the computational complexity. In [21] and [23], a block-adaptive LMS-type algorithm was proposed. However, additional BF orthogonalization or prewhitening had to be applied to ensure fast and reliable convergence. In [27], an SO solution was applied and demonstrated to achieve a similar performance as other state-of-the-art techniques. However, the CM was assumed fixed, and the computational cost of its recalculation was high. In [24]–[26] and [28], cascaded Hammerstein and Wiener structures were proposed. Such approaches typically have less free parameters, leading to solutions with lower associated complexity. However, they are best applied in scenarios where physical knowledge of the system under study is available, which is not always the case. Finally, works in [29] and [30] presented lookup table (LUT)-based approaches. The LUTs aimed at substituting the high-order polynomials in the DPD model, relaxing the overall complexity. Nonetheless, their modeling capabilities are somewhat limited, depending on the size of the LUTs. To enhance the performance, the LUT sizes can be increased to better describe the nonlinear characteristics, but this leads to higher memory requirements and slower convergence speeds.

In this article, we adopt a closed-loop MP structure in combination with a modified version of the SO learning to update the DPD model coefficients. We adopt the MP structure, in contrast to other more complex solutions, due to the relaxed linearity requirements of the mmW NR FR-2 systems [7], [14] and also because its reduced complexity makes it appealing for various DPD applications, such as the terminal PA linearization. The overall DPD structure is shown in Fig. 1, in the context of mmW active arrays.

The main technical contributions of this article are the development of several low-complexity solutions for estimating and manipulating the inverse CM (ICM) needed in SO learning algorithms. The ICM gathers statistical information of the DPD input data and the corresponding BF samples, and it is an important ingredient in the learning path to ensure good linearization performance and convergence. This is the most computationally heavy element in the SO learning rule and needs to be recalculated whenever the transmit signal type or its spectral characteristics are modified, to ensure optimal DPD execution. Specifically, we propose the following four methods:

1) method to efficiently modify the ICM by removing rows and columns for a smaller DPD parameterization;
2) method to efficiently modify the ICM when the input signal is frequency-shifted, reflecting, e.g., a change in resource allocation inside the channel bandwidth;
3) method to estimate the ICM from the autocorrelation function of the input data, assuming that it is a complex proper Gaussian distributed random signal.

4) Method to efficiently approximate the technique in 3) by utilizing the Bussgang decomposition.

Techniques 1) and 2) can be applied to modify a previously calculated ICM with low computational effort, while techniques 3) and 4) require only an estimate of the autocorrelation function of the input data to build the ICM. The SO learning algorithm in combination with the proposed methods allows for fast DPD coefficient estimation with minimal complexity in time-varying scenarios. Hence, on-chip implementations and continuous tracking of the DPD can become feasible.

Extensive over-the-air (OTA) RF measurement results at the 28 GHz mmW center frequency–the 5G NR band n257 [7]—are carried out and reported in order to test and validate the proposed solutions. All measurements feature a state-of-the-art 64-element active antenna array and 5G NR compliant OFDM waveforms, with signal bandwidths ranging from 200 to 400 MHz. In addition, further measurements with an aggregated signal bandwidth of 800 MHz are also conducted, with the aim of pushing the performance boundaries of the considered DPD system. Altogether, the obtained results, alongside with the complexity analyses, indicate very favorable performance–complexity tradeoffs of the proposed solutions when comparing against the state of the art. Finally, we note that although the experiments in this article consider mmW active array linearization, the developed solutions are applicable also in any more classical single-input–single-output DPD system and PA linearization context.

The rest of this article is organized as follows. Section II presents the adopted MP structure and also clarifies the assumptions taken for the mmW setup. Sections III and IV present the proposed low-complexity methods to estimate the ICM, needed for the SO learning rule. Complexity analyses and comparisons are then provided in Section V. Section VI presents the deployed 28 GHz measurement setup and the corresponding FR-2 RF measurement results and their analyses. Finally, conclusions are drawn in Section VII.

Notation Used in This Article: In this article, matrices are represented by capital boldface letters, e.g., $\mathbf{\Omega} \in \mathbb{C}^{M\times N}$. Ordinary transpose, Hermitian transpose, and complex conjugation are denoted by $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^*$, respectively. By default, vectors are complex-valued column vectors, presented with lowercase boldface letters, i.e., $\boldsymbol{\alpha} \in \mathbb{C}^{M\times 1} = [\alpha_1\ \alpha_2\ \cdots\ \alpha_M]^T$. In addition, the expected value, absolute value, floor, ceil, factorial, Hadamard product, and Kronecker product operators are written as $E\{\cdot\}$, $|\cdot|$, $\lfloor * \rfloor\cdot$, $\lceil * \rceil\cdot$, !, $\circ$, and $\otimes$, respectively.

II. Closed-Loop DPD System

In this section, the MP DPD model utilized in this work is described. Both the processing and learning paths are detailed, where the actual digital predistortion and the DPD coefficient estimation are executed. The MP model is adopted because it is a widely used model and is known to strike a good balance between linearization performance and computational complexity [1], [3], [5], [21], [23]. In addition, the DPD system builds on a closed-loop structure, where the DPD coefficients are estimated from the input signal, $x[n]$, and the observed learning signal, $y[n]$, [31] according to the notation followed in Fig. 1.

A. DPD Main Path

The complete closed-loop MP DPD scheme is presented in Fig. 1. Here, $x[n]$ is the original baseband signal, $x_{\text{DPD}}[n]$ is the signal after predistortion, and $y_1(t), y_2(t), \ldots, y_R(t)$ are the individual PA output signals to be transmitted. Following this notation and adopting the classical MP model structure, the input–output relation of the digital predistorter can be written as

$$x_{\text{DPD}}[n] = \sum_{m=0}^{M-1} \sum_{\substack{p=0 \\ p\ \text{odd}}}^{P} \alpha_{m,p} x[n-m]|x[n-m]|^{(p-1)} \tag{1}$$

where M denotes the number of memory taps considered in the model, P corresponds to the polynomial order, and $\alpha_{m,p}$ is the corresponding PA model coefficient for a specific memory tap and polynomial order. Following this configuration, the model has $C = \lceil (P/2) \rceil M$ coefficients. It is noted that while only odd orders are considered in the above model, there are works [32], [33], which have shown that the use of both odd and even orders can have certain benefits, such as the ability to reduce the overall polynomial order P for given linearization performance targets. The ICM estimation methods derived and described in this article focus on MP systems with odd orders, while the extension to cover also even-order polynomials is an interesting topic for future work in this area.

Next, in order to facilitate the closed-loop processing, the input–output relation obtained in (1) can be alternatively expressed with matrix notation as the complex multiplication of the MP BF matrix and the DPD coefficient vector. This expression reads

$$\mathbf{x}_{\text{DPD}} = \mathbf{\Omega}\boldsymbol{\alpha} \tag{2}$$

where $\boldsymbol{\alpha} \in \mathbb{C}^{C\times 1}$ contains the PA model coefficients and $\mathbf{\Omega} \in \mathbb{C}^{N\times C}$ is the original monomial BF matrix with N denoting the total number of the available input data samples. $\mathbf{\Omega}$ can be further expressed as

$$\mathbf{\Omega} = \big[\mathbf{x}_{1,0}\ \mathbf{x}_{1,1}\ \cdots\ \mathbf{x}_{1,M-1}\ \mathbf{x}_{3,0}\ \mathbf{x}_{3,1}\ \cdots\ \mathbf{x}_{3,M-1}\ \cdots\ \cdots\ \mathbf{x}_{P,0}\ \mathbf{x}_{P,1}\ \cdots\ \mathbf{x}_{P,M-1}\big] \tag{3}$$

with each vector $\mathbf{x}_{p,m} \in \mathbb{C}^{N\times 1}$ being defined as

$$\mathbf{x}_{p,m} = \begin{bmatrix} x[n-m]|x[n-m]|^{p-1} \\ x[n-m+1]|x[n-m+1]|^{p-1} \\ \vdots \\ x[n-m+N-1]|x[n-m+N-1]|^{p-1} \end{bmatrix}. \tag{4}$$

The order of the BF terms in (3) becomes important for the formulation taken in the methods presented in Sections III and IV. We finally note that in the first DPD iteration, the vector $\boldsymbol{\alpha}$ is usually initialized with a one in the first element and zeros in the rest (i.e., $\boldsymbol{\alpha} = [1\,0\cdots 0]^T$) such that only the unmodified linear term is passed through the predistorter.

B. DPD Learning Path

In closed-loop DPD systems, the observation signal $y[n]$ is nonlinear in the DPD parameters. Therefore, nonlinear LS

techniques need to be applied, such as those based on the GN algorithm [29], [34]. At block iteration k, these iterative algorithms build on the error vector $\mathbf{e}_k = \mathbf{y}_k - \mathbf{x}_k$ that is stacking the prevailing error signal samples of the current processing block of size K. In this article, we focus on the SO and the damped GN learning methods [21], [22], [29]. In Sections III and IV, the SO learning rule is studied in detail and combined with different methods to reduce the involved computational complexity, while the GN serves as a reference solution. To this end, the SO and the GN learning rules read

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{s}}\left(\mathbf{R}^*\right)^{-1}\boldsymbol{\Omega}_k^H\mathbf{e}_k \tag{5}$$

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{g}}\left(\boldsymbol{\Omega}_k^H\boldsymbol{\Omega}_k\right)^{-1}\boldsymbol{\Omega}_k^H\mathbf{e}_k \tag{6}$$

respectively, where μ_{s} and μ_{g} are the learning steps for the updates, $\boldsymbol{\Omega}_k \in \mathbb{C}^{K\times C}$ is the BF matrix corresponding to the DPD block iteration k, and $\mathbf{R} \in \mathbb{C}^{C\times C}$ is the CM of the input BF vector, defined as

$$\mathbf{R} = E\left\{\psi[n]\psi^H[n]\right\} \approx \frac{1}{N}\sum_{n=1}^{N}\psi[n]\psi^H[n]. \tag{7}$$

In above, $\psi[n]$ is the instantaneous input vector at time instant n (i.e., a row of $\boldsymbol{\Omega}$), and N is the total number of available samples over which $\mathbf{R}$ is estimated. If the input signal does not vary significantly within DPD iterations, the CM can be estimated off-line and then kept fixed, avoiding repetitive online calculations. For later purposes, we also already define the ordinary autocorrelation function of the DPD input signal $x[n]$, as well as its sample estimate, as

$$R_X(\tau) = E\left\{x[n]x^*[n-\tau]\right\} \approx \frac{1}{N}\sum_{n=1}^{N}x[n]x^*[n-\tau]. \tag{8}$$

In general, it is important to note that the term $(\boldsymbol{\Omega}_k^H\boldsymbol{\Omega}_k)^{-1}$ in the GN learning rule essentially calculates the (conjugate of the) ICM at every DPD iteration. This provides an accurate description of the statistics of the input signal at every iteration but also involves heavy computational complexity. For this reason, the GN method is considered as the baseline high-performance solution against which the linearization performance is compared. At the same time, the SO approach aims at drastically reducing the computational complexity by obtaining the ICM with different alternatives, described in Sections III and IV, while still achieving a linearization performance close to GN. We further note that the learning solutions presented in (5) and (6) are generic learning approaches for linear-in-parameters models and can also be applied to other models, such as GMP [3], Volterra DDR [35], [36], or even LUT-based solutions [29], [37].

Finally, it is noted that one common alternative in the DPD developments is to deploy BF orthogonalization or whitening [21], [23]. Such an approach can also facilitate utilizing low-complexity learning algorithms, even block-LMS [22]. However, the BF whitening also implies increased DPD main path complexity—even more so in case of dynamic input signals, implying that also the whitening matrix should be recalculated. Hence, in this work, we do not consider BF whitening in the DPD main path.

C. Observation Receiver Configuration in mmW Active Array Systems

In the context of mmW phased-array systems, there exist various alternatives for arranging the ORX in order to obtain the combined learning signal—$y[n]$ in Fig. 1—used for DPD parameter learning. One particular approach is the hardware-based method, in which each PA output signal is coupled, phase-aligned, and combined in hardware [11], [12], [38]. Another plausible solution is to configure a separate ORX, which captures the combined OTA received signal and feeds it back to the DPD systems for DPD learning [13], [39]–[41]. These two alternatives are presented, as plausible solutions, in Fig. 1. Both of these alternatives mimic the far-field combined signal at a distant receiver, in case of line-of-sight propagation.

It is important to clarify that the DPD learning solutions and methods presented within this article do not depend on the actual way of obtaining the combined learning signal used for parameter estimation. It is noted, however, that the hardware-based approach does not suffer from the common OTA-related challenges, such as ORX misalignment and positioning challenges, environmental dependencies, and beam dependence of radiated nonlinear distortion. We also note that there exist several works in the literature [39]–[41], where the OTA ORX beam misalignment challenges are studied, proposing different solutions to provide the far-field signal in the main beam direction. Furthermore, our adopted mmW measurement setup, featured in Section VI, aims at mimicking the hardware-based feedback system by carefully aligning a horn antenna in the main beam, acting as the OTA ORX.

III. Low-Complexity Methods for Modifying the ICM

If the input signal to the SO algorithm remains static, the ICM can be precalculated off-line, saving complexity in the DPD learning path. However, whenever the input signal changes, the ICM needs to be estimated again in order to provide optimal linearization performance. The process of calculating the ICM—or, alternatively, the term $(\boldsymbol{\Omega}_k^H\boldsymbol{\Omega}_k)^{-1}$ in the GN learning rule (6)—is computationally heavy, as it involves calculating the inverse of the product between the BF matrix and its Hermitian transpose. The size of the BF matrix, $\boldsymbol{\Omega}_k$, is $K \times C$, which is typically large in DPD implementations. To ease this calculation, we propose in Sections III-A and III-B two methods for modifying the ICM when the input signal or the parameterization of the DPD is changed. We finally note that these solutions are applicable to any DPD models that are linear in the parameters, not just the MP.

A. Reducing the Dimensionality of the ICM

The first considered method avoids the need of recalculating the ICM if a simpler parameterization is desired in the DPD system. A simpler parameterization might be adopted when lowering the transmit power or when changing from a high-order modulation to a lower order modulation with less stringent EVM requirements. Furthermore, in mmW active array linearization, where the severity of the nonlinear

distortions varies as a function of the beam direction, such a technique may be particularly useful in optimizing the power consumption of the transmitter.

To this end, we first generate a generic ICM obtained with a large DPD parameterization. If a smaller polynomial order or memory depth is desired, the proposed method then removes the appropriate rows and columns from the generic ICM and obtains a new ICM submatrix, which corresponds to the new chosen parameterization. The new submatrix can then be used in the SO learning equation to estimate the new DPD coefficients.

The proposed method can be described as follows. Let us consider a generic ICM, $\tilde{\mathbf{Q}}^{-1} = (\tilde{\mathbf{R}}^*)^{-1} \in \mathbb{C}^{\tilde{C}\times\tilde{C}} = (\tilde{\mathbf{\Omega}}^H\tilde{\mathbf{\Omega}})^{-1}$, obtained from the complete set of BFs, $\tilde{\mathbf{\Omega}}$. This matrix is known and it is used to find the reduced ICM subset $\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1} \in \mathbb{C}^{C\times C} = (\mathbf{\Omega}^H\mathbf{\Omega})^{-1}$, as described before. The BF matrices, $\tilde{\mathbf{\Omega}}$ and $\mathbf{\Omega}$, are mutually related as

$$\tilde{\mathbf{\Omega}} = \begin{pmatrix}\mathbf{\Omega} & \mathbf{v}\end{pmatrix} \tag{10}$$

where $\mathbf{v} \in \mathbb{C}^{N\times 1}$ represents a column in $\tilde{\mathbf{\Omega}}$ that needs to be removed. The generic CM can be now written as

$$\tilde{\mathbf{Q}} = \tilde{\mathbf{\Omega}}^H\tilde{\mathbf{\Omega}} = \begin{pmatrix}\mathbf{\Omega}^H \\ \mathbf{v}^H\end{pmatrix}\begin{pmatrix}\mathbf{\Omega} & \mathbf{v}\end{pmatrix} = \begin{pmatrix}\mathbf{\Omega}^H\mathbf{\Omega} & \mathbf{\Omega}^H\mathbf{v} \\ \mathbf{v}^H\mathbf{\Omega} & \mathbf{v}^H\mathbf{v}\end{pmatrix} \tag{11}$$

and its block inverse can be directly obtained by using the Schur complement [42] of (11) as

$$\tilde{\mathbf{Q}}^{-1} = \left(\begin{array}{c|c} \mathbf{A}^{-1} & \mathbf{B}^{-1} \\ \hline \mathbf{C}^{-1} & \mathbf{D}^{-1} \end{array}\right) \tag{12}$$

where

$$\mathbf{A}^{-1} = \left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1} + \mathbf{D}^{-1}\left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^H\mathbf{v}\mathbf{v}^H\mathbf{\Omega}\left(\left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1}\right)^H \tag{13}$$

$$\mathbf{B}^{-1} = -\mathbf{D}^{-1}\left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^H\mathbf{v}, \tag{14}$$

$$\mathbf{C}^{-1} = -\mathbf{D}^{-1}\mathbf{v}^H\mathbf{\Omega}\left(\left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1}\right)^H \tag{15}$$

$$\mathbf{D}^{-1} = \left(\mathbf{v}^H\mathbf{v} - \mathbf{v}^H\mathbf{\Omega}\left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1}\mathbf{\Omega}^H\mathbf{v}\right)^{-1}. \tag{16}$$

Finally, the system of linear equations in (13) can be directly solved to obtain $(\mathbf{\Omega}^H\mathbf{\Omega})^{-1}$, which corresponds to the desired ICM subset, expressed as

$$\mathbf{Q}^{-1} = \left(\mathbf{R}^*\right)^{-1} = \left(\mathbf{\Omega}^H\mathbf{\Omega}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1}\left(\frac{\mathbf{B}^{-1}}{\mathbf{D}^{-1}}\right)^H. \tag{17}$$

We note that removing one column from the BF matrix, $\tilde{\mathbf{\Omega}}$, corresponds to removing one row and one column from the ICM, $\tilde{\mathbf{Q}}^{-1}$. The proposed method requires moving the row and column to be removed to the end of the matrix and can be executed iteratively to cut out more than one row/column. In Algorithm 1, we present an example pseudocode that shows how to remove l rows and columns from $\tilde{\mathbf{Q}}^{-1} \in \mathbb{C}^{\tilde{C}\times\tilde{C}}$, obtaining the new ICM submatrix $\mathbf{Q}^{-1} \in \mathbb{C}^{(\tilde{C}-l)\times(\tilde{C}-l)}$.

Algorithm 1 Pseudocode showing how l rows and columns can be removed from $\tilde{\mathbf{Q}}^{-1}$ to obtain the subset $\mathbf{Q}^{-1}$. The generic ICM, given as input data, is conceptually presented in (12)

Data: $\tilde{\mathbf{\Omega}}$, $\tilde{\mathbf{Q}}^{-1}$
Result: $\mathbf{Q}^{-1}$
Initialize: $\mathbf{Q}^{-1} \in \mathbb{R}^{C\times C} = \mathbf{I}$,
$\mathbf{A}^{-1} \in \mathbb{R}^{(\tilde{C}-1)\times(\tilde{C}-1)} = \mathbf{0}$,
$\mathbf{B}^{-1} \in \mathbb{R}^{(\tilde{C}-1)\times 1} = \mathbf{0}$,
$\mathbf{D}^{-1} = 0$;
for $i = 1, 2, \ldots, l-1, l$, **do**
  Permute row/column to the end of $\tilde{\mathbf{Q}}^{-1}$;
  $\mathbf{A}^{-1} = \tilde{\mathbf{Q}}^{-1}(1:(\tilde{C}-i),\ 1:(\tilde{C}-i))$;
  $\mathbf{D}^{-1} = \tilde{\mathbf{Q}}^{-1}(\tilde{C}-i+1, \tilde{C}-i+1)$;
  $\mathbf{B}^{-1} = \tilde{\mathbf{Q}}^{-1}(1:(\tilde{C}-i), \tilde{C}-i+1)$;
  $\mathbf{Q}^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1}(\frac{\mathbf{B}^{-1}}{\mathbf{D}^{-1}})^H$;
end

B. Frequency Shifting the ICM

In the second considered scenario, we study how to avoid the ICM recalculation when the DPD input signal is digitally frequency shifted. In such a case, the spectral components of the signal are modified, and thus, its CM and ICM will no longer be the same. In order to provide accurate linearization performance, the new ICM needs to be estimated again by some means. To that end, we propose a novel method that shifts the original ICM, denoted as $\tilde{\mathbf{Q}}^{-1} = (\tilde{\mathbf{R}}^*)^{-1}$, in the frequency domain to obtain the resulting shifted ICM, denoted as $\mathbf{Q}^{-1} = (\mathbf{R}^*)^{-1}$. The computational complexity of this method is very low compared to recalculating the whole ICM again, as shown in Section V. A good example use case of such frequency shifting is when the uplink spectral allocation of a terminal/UE transmitter changes within a given channel bandwidth.

The proposed method is constructed as follows. Let us denote the original unshifted digital input signal as $\tilde{x}[n]$ and the corresponding shifted signal as $x[n]$. A frequency shift of Δf, in Hz, can be mathematically expressed in the time domain as the multiplication with a sampled complex exponential, and thus, the shifted digital signal reads

$$x[n] = e^{j\Delta\omega n}\tilde{x}[n] \tag{19}$$

where $\Delta\omega = (2\pi\Delta f)/f_s$ denotes the normalized angular frequency shift, while f_s refers to the sample rate. The next step is to obtain the shifted BF matrix, $\mathbf{\Omega}$, exclusively as a function of the unshifted signal $\tilde{x}[n]$ and the normalized frequency shift $\Delta\omega$. This can be done by substituting (19) into (3) and (4). For simplicity, we just present the first row of the resulting $\mathbf{\Omega}$, shown in (9) at the bottom of the next page, considering that $\left|e^{j\Delta\omega n}\right| = 1$. Due to the formulation taken in (9), $\mathbf{Q}$ can also be expressed as a function of $\tilde{\mathbf{Q}}$ and a multiplying factor that depends on the frequency shift $\Delta\omega$. This factor appears in the BF terms with different delays, caused by the distinct memory terms appearing in the BF matrix. Specifically, we observe 0-sample-delay terms,

1-sample-delay terms, and, in general, m-sample-delay terms, appearing in the CM, expressed as

$$\begin{aligned}
0\text{-sample-delay: } \mathbf{Q}(i,j) &= e^{-j\Delta\omega n}e^{j\Delta\omega n}\tilde{\mathbf{Q}}(i,j) = \tilde{\mathbf{Q}}(i,j)\\
1\text{-sample-delay: } \mathbf{Q}(i,j) &= e^{-j\Delta\omega n}e^{j\Delta\omega(n-1)}\tilde{\mathbf{Q}}(i,j)\\
&= e^{-j\Delta\omega 1}\tilde{\mathbf{Q}}(i,j)\\
&\vdots\\
m\text{-sample-delay: } \mathbf{Q}(i,j) &= \cdots = e^{-j\Delta\omega m}\tilde{\mathbf{Q}}(i,j). \qquad (20)
\end{aligned}$$

The CM is generally a Hermitian–Toeplitz matrix, and the way how the different sample delay terms are scattered through it can also be seen from (28), where the value of τ within the autocorrelation function $R_X(\tau)$, defined formally in (8), indicates the corresponding delay. Hence, the zero-delay samples correspond to the main diagonal of the $M \times M$ submatrices within the CM, the one-delay samples correspond to the second diagonal term of these submatrices, and so on. From this structure, it is then relatively easy to see that the frequency-shifted CM can be written as

$$\mathbf{Q} = \mathbf{P}^H\tilde{\mathbf{Q}}\mathbf{P} \qquad (21)$$

where the diagonal matrix $\mathbf{P} \in \mathbb{C}^{C\times C}$ applies the right shift weights for each column of $\tilde{\mathbf{Q}}$, and it is defined as

$$\mathbf{P} = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & e^{j\Delta\omega m} & 0 & \cdots & 0\\ 0 & \cdots & 0 & 1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & e^{j\Delta\omega m} \end{bmatrix}. \qquad (22)$$

Defining the CM as in (21) will allow for applying simple yet efficient matrix inverse properties. As a result, the shifted ICM can be directly obtained from (21) as

$$\mathbf{Q}^{-1} = \left(\mathbf{R}^*\right)^{-1} = \mathbf{P}^{-1}\tilde{\mathbf{Q}}^{-1}\mathbf{P}^{H(-1)}. \qquad (23)$$

Note that only the original ICM, $\tilde{\mathbf{R}}^{-1}$, and the frequency shift, $\Delta\omega$, are needed to obtain the final shifted ICM, $\mathbf{R}^{-1}$.

IV. Low-Complexity ICM Estimation for Gaussian Signals

This section presents two novel methods to estimate the ICM in a computationally efficient manner. These methods build on the autocorrelation function of the input signal and provide a set of precomputed coefficients that can be utilized to easily estimate the ICM. The first method is an application of Reed's theorem [43] for calculating the CM, while the second technique utilizes the Bussgang theorem and other stochastic analysis to provide an approximate CM. In the latter case, the resulting CM can be shown to consist of a specific subblock structure expressible as a Kronecker product, allowing for efficient inversion. We note that, since we consider complex Gaussian process theory, we assume the Gaussian distributed input signals in this section. OFDM signals are known to converge toward Gaussian, as the number of subcarriers is large [44]. This claim is further substantiated in Section VI.

A. ICM Estimation With the Autocorrelation Function

The first proposed method utilizes the moment theorem for complex-circular Gaussian signals presented in [43]. We specifically aim at expressing the CM as a function of the autocorrelation function of the input signal. To that end, the autocorrelation function is first estimated over the desired span with (8). Reed's theorem states that all odd-order central product moments are zero, while all even-order moments can

$$\begin{aligned}
\mathbf{\Omega} = \big(&x[n]e^{j\Delta\omega n} \ \cdots \ x[n]|x[n]|^{(P-1)}e^{j\Delta\omega n} \quad x[n-1]e^{j\Delta\omega(n-1)} \ \cdots \ x[n-1]|x[n-1]|^{(P-1)}e^{j\Delta\omega(n-1)} \ \cdots\\
&\cdots \ x[n-M+1]e^{j\Delta\omega(n-M+1)} \ \cdots \ x[n-M+1]|x[n-M+1]|^{(P-1)}e^{j\Delta\omega(n-M+1)}\big) \qquad (9)
\end{aligned}$$

$$\begin{aligned}
E\{x[n]x^*[n-1]|x[n-1]|^2\} &= E\{x^*[n-1]x[n]\}E\{x^*[n-1]x[n-1]\} + E\{x^*[n-1]x[n]\}E\{x^*[n-1]x[n-1]\}\\
&= 2R_X(\tau=0)R_X(\tau=1),\\
E\{x[n-1]x^*[n]|x[n]|^2\} &= E\{x^*[n]x[n]\}E\{x^*[n]x[n-1]\} + E\{x^*[n]x[n]\}E\{x^*[n]x[n-1]\}\\
&= 2R_X(\tau=0)R_X^*(\tau=1),\\
E\{x[n]x^*[n-1]|x[n]|^2|x[n-1]|^2\} &= E\{x^*[n]x[n]\}E\{x^*[n-1]x[n]\}E\{x^*[n-1]x[n-1]\}\\
&\quad + E\{x^*[n]x[n]\}E\{x^*[n-1]x[n]\}E\{x^*[n-1]x[n-1]\}\\
&\quad + E\{x^*[n-1]x[n]\}E\{x^*[n]x[n]\}E\{x^*[n-1]x[n-1]\}\\
&\quad + E\{x^*[n-1]x[n]\}E\{x^*[n-1]x[n]\}E\{x^*[n]x[n-1]\}\\
&\quad + E\{x^*[n-1]x[n]\}E\{x^*[n]x[n]\}E\{x^*[n-1]x[n-1]\}\\
&\quad + E\{x^*[n-1]x[n]\}E\{x^*[n-1]x[n]\}E\{x^*[n]x[n-1]\}\\
&= 4R_X^2(\tau=0)R_X(\tau=1) + 2R_X^2(\tau=1)R_X^*(\tau=1),\\
E\{x[n-1]x^*[n]|x[n]|^2|x[n-1]|^2\} &= \quad \cdots \quad = 4R_X^2(\tau=0)R_X^*(\tau=1) + 2\left(R_X^*(\tau=1)\right)^2R_X(\tau=1) \qquad (18)
\end{aligned}$$

be obtained through sums of products of the second-order moments, expressed in the general case as [43]

$$E\{x_0^* x_1^* \cdots x_{k-1}^* x_0 x_1 \cdots x_{k-1}\} = \sum_{\zeta} E\{x_\zeta^* x_0\} E\{x_\zeta^* x_1\} \cdots E\{x_\zeta^* x_{k-1}\}. \quad (25)$$

Here, for notational simplicity, $x_m = x[n-m]$, and ζ constitutes a permutation of the set $\{0, 1, \ldots, (k-1)\}$. Noting that all terms in the CM have an even order, and following the specific BF order assumed in (3), each term in the CM can be obtained by applying (25).

For clarity, an example of the application of Reed's theorem to the fourth- and sixth-order moments is presented in (18), as shown at the bottom of the previous page, while other higher order terms can be obtained similarly. The definition of the autocorrelation function in (8) has been used to arrive at the last expressions in (18). It is noted that the expression presented in (25) allows to obtain the higher order terms recursively. It is also noted that the specific terms of the form $E\{|x[n]|^{2p}\}$ can be alternatively obtained as

$$E\{|x[n]|^{2p}\} = p! E\{|x[n]|^2\}^p = p! \sigma_x^{2p} \quad (26)$$

which may simplify their calculation.

Once Reed's theorem has been applied to each individual term in the CM, we can express the CM with a specific sub-block matrix structure. Specifically, the CM can be expressed with a $\lceil (P/2) \rceil \times \lceil (P/2) \rceil$ submatrix structure, each of size $M \times M$, written as presented in (24), as shown at the bottom of the page. Here, the submatrix $\mathbf{R}_{k,l}$ is in turn defined as

$$\mathbf{R}_{k,l} = E\left\{ \begin{bmatrix} \psi_k[n]\psi_l^*[n] & \cdots & \psi_k[n]\psi_{l+M-1}^*[n] \\ \vdots & \ddots & \vdots \\ \psi_{k+M-1}[n]\psi_l^*[n] & \cdots & \psi_{k+M-1}[n]\psi_{l+M-1}^*[n] \end{bmatrix} \right\} \quad (27)$$

where the subscript in $\psi[n]$ indicates the corresponding element within the instantaneous input vector. Due to the notation utilized in (24) and (27), such matrices can accommodate any polynomial order, P, and memory depth, M, chosen in the DPD system. Note that only the calculation of the upper triangular submatrices in $\mathbf{R}$ is required to build the complete CM. The CM is then inverted to generate the resulting ICM, thereon used in the SO learning equation.

B. ICM Approximation With Bussgang Coefficients

In the second proposed method, we aim at simplifying the previous method by using Bussgang's decomposition. This theorem states that the cross correlation of a Gaussian signal $x[n]$ and a signal $y[n] = f(x[n])$ that has passed through an instantaneous nonlinear function $f(\cdot)$ can be expressed as the product of the autocorrelation function of $x[n]$ and a scaling constant. Formally, this is expressed as

$$R_{XY}(\tau) = \xi R_X(\tau) \quad (29)$$

where

$$R_{XY}(\tau) = E\{x[n] y^*[n-\tau]\} \quad (30)$$

is the definition of the cross correlation function, while the autocorrelation function $R_X(\tau)$ reads as in (8). The Bussgang coefficient, ξ, is obtained through the complex proper Gaussian probability distribution function (PDF) as

$$\xi = \frac{1}{\pi \sigma_x^4} \int_{-\infty}^{\infty} x^* f(x) e^{-\frac{|x|^2}{\sigma_x^2}} dx \quad (31)$$

where $f(\cdot)$ represents the nonlinear amplitude distortion and σ_x^2 is the variance of $x[n]$.

In general, the second-order terms appearing in the CM, on its first row, can be expressed directly as a function of $R_X(\tau)$ as

$$\mathbf{R}(1,1) = R_X(\tau = 0) \quad (32)$$
$$\mathbf{R}(1,2) = R_X(\tau = 1) \quad (33)$$
$$\vdots$$
$$\mathbf{R}(1,M) = R_X(\tau = M). \quad (34)$$

The higher order terms appearing in the CM (i.e., $x[n]|x[n]|^p$, $p = 2, 4, \ldots$) can be seen as nonlinear functions of $x[n]$ and can thus be expressed as a function of the autocorrelation according to the Bussgang theorem [45].

Using (29), we can now express all the remaining high-order terms in the CM uniquely as a function of $R_X(\tau)$ as

$$E\{x[n]x^*[n]|x[n]|^2\} = \xi_1 R_X(\tau = 0) \quad (35)$$
$$E\{x[n]x^*[n-1]|x[n-1]|^2\} = \xi_1 R_X(\tau = 1) \quad (36)$$
$$E\{x[n]x^*[n]|x[n]|^4\} = \xi_2 R_X(\tau = 0) \quad (37)$$
$$E\{x[n]x^*[n-1]|x[n-1]|^4\} = \xi_2 R_X(\tau = 1) \quad (38)$$
$$\vdots$$

where the coefficients $\xi_1, \xi_2, \ldots$ can be obtained using (31). By having all the terms as a function of $R_X(\tau)$ and the Bussgang coefficients, the complete CM can be expressed through them. A concrete example is presented in (28), as shown at the bottom of the next page, which is showing the CM structure in the specific case of $P = 3$ and $M = 1$ while also assuming that $\xi_0 = 1$. Each matrix element, which depends on the input data signal, can be then obtained by using the explicit expressions shown in (7) and (31).

We next note that the order of the BFs in the BF matrix becomes important to replicate this particular structure. By further studying (28), we realize that it can be formulated with a

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{1,1} & \mathbf{R}_{1,M+1} & \cdots & \mathbf{R}_{1,(\lceil \frac{P}{2} \rceil - 1)M+1} \\ \mathbf{R}_{1,M+1} & \mathbf{R}_{M+1,M+1} & \cdots & \mathbf{R}_{M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{1,(\lceil \frac{P}{2} \rceil - 1)M+1} & \mathbf{R}_{M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} & \cdots & \mathbf{R}_{(\lceil \frac{P}{2} \rceil - 1)M+1,(\lceil \frac{P}{2} \rceil - 1)M+1} \end{bmatrix} \quad (24)$$

specific subblock matrix notation that will allow for efficient Kronecker inversion. The CM can be alternatively simplified in $M \times M$ subblocks, which are weighted with a different Bussgang coefficient and are propagated through $\mathbf{R}$ as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} \\ \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{R}_3 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 1} \\ \mathbf{R}_2 & \mathbf{R}_3 & \mathbf{R}_4 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 1} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor + 2} & \cdots & \mathbf{R}_{2\lfloor \frac{P}{2} \rfloor} \end{bmatrix} \tag{39}$$

where the subscripts indicate the corresponding Bussgang coefficients. In addition, the submatrix $\mathbf{R}_k \in \mathbb{C}^{M \times M}$ reads

$$\mathbf{R}_k = \xi_k \begin{bmatrix} R_X(\tau = 0) & R_X(\tau = 1) & \cdots & R_X(\tau = M) \\ R_X^*(\tau = 1) & R_X(\tau = 0) & \cdots & R_X(\tau = M - 1) \\ \vdots & \vdots & \ddots & \vdots \\ R_X^*(\tau = M) & R_X^*(\tau = M - 1) & \cdots & R_X(\tau = 0) \end{bmatrix}. \tag{40}$$

Note that only the submatrices appearing on the first row and the last column of (39) need to be calculated to build the whole CM.

Finally, due to the well-structured formulation in (39), the CM can be efficiently inverted using different matrix inversion methods. One possible solution is to use block-recursive matrix inverse algorithms, such as the one in [46], which poses drastically reduced complexity when compared to ordinary matrix inversion. An alternative and yet more efficient solution is to use the Kronecker inversion [47], [48]. To this end, we first note that the CM can be equivalently expressed with the Kronecker product as

$$\mathbf{R} = \Xi \otimes \mathbf{R}_0 \tag{41}$$

where $\Xi \in \mathbb{R}^{\lceil (P/2) \rceil \times \lceil (P/2) \rceil}$ contains the complete set of Bussgang coefficients. Then, the ICM is directly obtained as the inverse of its elements as

$$\mathbf{R}^{-1} = \Xi^{-1} \otimes \mathbf{R}_0^{-1}. \tag{42}$$

The Kronecker inversion further reduces the computational complexity of inverting $\mathbf{R}$, as demonstrated in Section V.

V. Complexity Analysis and Comparison

In this section, we present the computational complexity involved in the proposed DPD and ICM estimation solutions. The processing complexity is divided in three different stages—the DPD main path (Section II-A), the DPD learning path (Section II-B), and the ICM-related calculations (Sections III and IV). All the analyses are carried out in terms of real multiplications per K-sized DPD iteration (i.e., for one block iteration with data size of K samples). Multiplications constitute, in general, the most resource-intensive operations in digital signal processing (DSP) implementations, while additions can be considered to be essentially free [15], [49]. In this study, it is assumed that one complex-complex multiplication is calculated with four real multiplications, and one real-complex multiplication is calculated with two real multiplications. In addition, the matrix inversion of an $m \times m$ matrix is assumed to cost 4 m^3 real multiplications [42], [46].

As noted, the first considered stage is the DPD main path, where the actual predistortion is applied. Here, the complete BF matrix [i.e., $\mathbf{\Omega}$ in (3)] is first calculated from the input signal $x[n]$, assuming its recursive calculation through previously obtained values. Later, the actual predistortion processing is carried out as in (2). The second considered stage is the DPD learning path, where the closed-loop learning equations are applied to update the DPD coefficients. Here, we consider both the reference GN algorithm and the SO learning approach. In this analysis, in case of SO learning, we express the learning complexity first for a given ICM while noting the ICM calculation complexity then separately. The GN approach, in turn, needs to calculate the ICM estimate $(\mathbf{\Omega}_k^H \mathbf{\Omega}_k)^{-1}$ in every DPD iteration—by its very definition. Finally, a complexity assessment and comparison between the classical ICM calculation and the proposed ICM estimation methods are provided. The classical ICM estimation complexity directly results from the calculation of $(\mathbf{\Omega}^H \mathbf{\Omega})^{-1}$. The complexity of the proposed ICM estimation methods, in turn, is determined following the exact processing principles presented in Sections III and IV.

Table I gathers the obtained complexity analysis results for the considered methods. The table presents the required real multiplications in each stage of the DPD solution, in a general form as a function of the modeling parameters. In addition, the final column presents a specific numerical example when $P = 9$, $M = 4$, $C = 20$, $l = 5$, and $K = 20\,000$, in order to provide a concrete and representative example of the involved complexities. Such DPD parameterization is similar to the one chosen to perform the RF experiments and validation in Section VI. From the numerical example, it can be clearly observed how the computational complexity is reduced against the reference GN method when utilizing the SO learning solution in combination with the proposed ICM estimation methods. In all cases, the learning complexity is drastically reduced while still achieving a similar linearization performance, as demonstrated in Section VI. The quantitative results also demonstrate the very large reduction in the needed multiplications, achievable through the proposed ICM estimation methods, compared to the direct ICM calculation.

$$\mathbf{R} = \begin{bmatrix} \xi_0 R_X(\tau = 0) & \xi_0 R_X(\tau = 1) & \xi_1 R_X(\tau = 0) & \xi_1 R_X(\tau = 1) \\ \xi_0 R_X^*(\tau = 1) & \xi_0 R_X(\tau = 0) & \xi_1 R_X^*(\tau = 1) & \xi_1 R_X(\tau = 0) \\ \xi_1 R_X(\tau = 0) & \xi_1 R_X(\tau = 1) & \xi_2 R_X(\tau = 0) & \xi_2 R_X(\tau = 1) \\ \xi_1 R_X^*(\tau = 1) & \xi_1 R_X(\tau = 0) & \xi_2 R_X^*(\tau = 1) & \xi_2 R_X(\tau = 0) \end{bmatrix} \tag{28}$$

TABLE I
COMPUTATIONAL COMPLEXITIES OF THE DPD MAIN PATH, DPD LEARNING PATH, AND THE VARIOUS ICM ESTIMATION METHODS

Operation		**Amount of real multiplications**	**Numerical example with** $P = 9$, $M = 4$, $C = 20$ $l = 5$, $K = 20,000$
DPD main path	**BF generation**	$2K(\lfloor \frac{P}{2} \rfloor + 1)$	200,000
	Digital predistortion	$4KC$	1,600,000
	Total	$2K(\lfloor \frac{P}{2} \rfloor + 2C + 1)$	1,800,000
DPD learning path	**Damped Gauss-Newton**	$4C(C^2 + CK + C + K + \frac{1}{2})$	33,633,640
	Self-orthogonalization	$4C(K + C + \frac{1}{2})$	1,601,640
Other: ICM calculation (only when required)	**Classical ICM calculation**	$4C^2(C + K)$	32,032,000
	Dimensionality red. ICM (Sec. III-A)	$4lC(C - 1)$	7,600
	Frequency shifting ICM (Sec. III-B)	$(C^2 - C)/2$	190
	Autocorr. ICM est. (Sec. IV-A)	$(\lfloor \frac{P}{2} \rfloor^2 + 3\lfloor \frac{P}{2} \rfloor + 2)3M^2 + 4(C^3(M + 1)(K + 1))$	433,460
	Bussgang's ICM est. (Sec. IV-B)	$M(4M^2 + 6M\lceil \frac{P}{2} \rceil - 3M + 4K + 4) + 4(\lceil \frac{P}{2} \rceil + K + 1)$	401,208

For reference and comparison purposes, we also address the complexity of a classical MP model with BF prewhitening or orthogonalization while considering block-LMS as the DPD coefficient update rule. With block-LMS, BF prewhitening becomes a necessary ingredient stemming from the large eigenvalue spread of the input signal CM. This process ensures that reasonably fast and stable learning convergence can be obtained by the simple LMS algorithm, which essentially uses a diagonal matrix to approximate the ICM [22]. The BF prewhitening is done in the DPD main path and is assumed here to build on the well-known Cholesky decomposition-based method [23], [42], which provides an upper triangular orthogonalization matrix. The overall main path complexity of such an approach is then the one presented on the fourth row of Table I plus the extra cost of the prewhitening stage that is a matrix–matrix product. Due to the triangular orthogonalization matrix, the extra complexity is of the form $4K(C(C + 1)/2)$. In the numerical example case of Table I, the corresponding overall main path complexity is 18.6M real multiplications, reflecting a substantial increase in comparison to the methods that do not require prewhitening. On the other hand, the cost of updating the DPD coefficients in the block-LMS learning rule [50] is given by $4CK + 2C$, which leads to 1 600 040 real multiplications—a number that is of the same complexity order as the SO solution when combined with any of the proposed ICM estimation methods. This analysis thus shows that the main path complexity increases substantially when BF orthogonalization is deployed, while the corresponding decrease in the learning complexity through using block-LMS is fairly minor when the ICM calculations in the reference SO solutions build on the proposed methods. In addition, we note that in these complexity calculations, the prewhitening matrix has been considered to be precalculated off-line. However, when considering DPD applications for example in terminal transmitters, where the transmitted waveforms are very dynamic (the resource block allocation and/or output power can change on a per-slot basis), such an assumption may not be feasible. This, in turn, means that in such dynamic scenarios, also the prewhitening matrix may need to be recalculated online during the learning procedures, further increasing the corresponding online processing burden.

Finally, based on the results in Table I, we separate three possible alternatives regarding the SO learning approach in combination with the ICM estimation.

1) The SO learning rule is applied in a dynamic waveform scenario using classical ICM calculation based on the sample estimate of the CM in (7). The ICM is thus calculated each time the waveform changes. If this happens frequently, the complexity approaches that of the GN.
2) The SO learning rule is applied in a dynamic waveform scenario using the proposed ICM estimation methods. Computational complexity may be drastically reduced compared to the previous when waveform changes are frequent.
3) The SO learning rule is applied in a static waveform scenario. In this case, the ICM is fixed and only needs to be computed once. Thus, the overall complexity is the smallest.

VI. RF MEASUREMENTS

In order to test, verify, and validate the proposed methods and algorithms, an extensive set of RF measurements is provided, in the context of an FR-2 mmW device. The deployed system features a state-of-the-art 28 GHz 64-element active phased array, acting as the nonlinear element which the proposed methods aim at linearizing. The measured 1-dB compression point of the array is +41 dBm, and the operating FR-2 band is 5G NR band n257. Once the measurement results are obtained, they are analyzed together with the quantitative complexity analysis results provided in Section V, to assess the performance–complexity tradeoffs of the proposed solutions.

In the context of mmW measurements with multiple-element phased arrays, several issues are to be noted. First, in an R-antenna array, there are also R parallel PAs, one per antenna unit. Each of these amplifiers will basically have

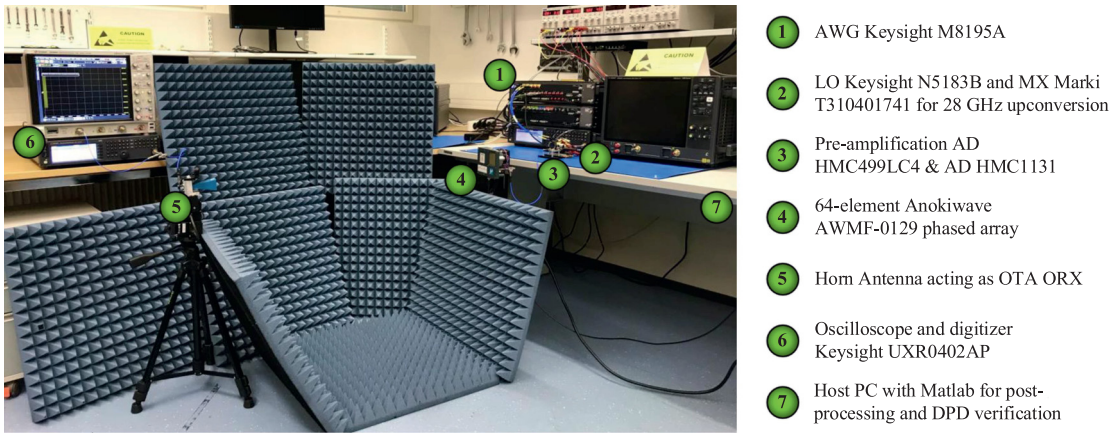


Fig. 2. 5G NR mmW/FR-2 OTA measurement setup utilized to carry out the experiments at 28 GHz.

unique nonlinear characteristics, and hence, the predistorter can only guarantee an efficient linearization performance in the main beam direction. In the rest of the angles, the intrinsic beampattern of the array suppresses the OoB nonlinear distortion [11]. Another concern is the load modulation of the PAs, occurring due to coupling between the antenna elements [8]. This makes the effective nonlinear characteristics of the antenna array beam-dependent, thus tying the DPD solution to the beam direction. Consequently, linearization methods must take considered. This issue can be tackled with, for instance, real-time DPD tracking, which is able to estimate the DPD coefficients as the beam is steered. Since the parameter estimation needs to be done in real time, it is crucial to explore ways of minimizing the involved computational complexity while at the same time being able to provide the needed linearization performance.

In this work, the DPD performance is evaluated through the well-known EVM [15], normalized mean square error (NMSE) [15], and the TRP-ACLR metrics since an OTA system is considered [7]. The TRP-ACLR measures the OoB performance by computing the ratio between the filtered mean power centered on the assigned channel frequency and the filtered mean power centered on an adjacent channel frequency, measured by integrating the powers over the whole beamspace while keeping the beamforming angle fixed [7], [8].

A. 28 GHz Active Array Experimental Setup

The complete 28 GHz measurement setup is shown in Fig. 2. The experimental setup is configured as follows. First, a Keysight M8195A arbitrary waveform generator (AWG) is deployed to output the modulated I/Q waveform at 2.5 GHz intermediate frequency (IF). Then, a Keysight N5183B signal generator provides the LO signal at 24.5 GHz that together with a Marki Microwave T310401741 mixer further upconverts the signal to the 28 GHz band. A Marki Microwave FB3300 bandpass filter (BPF) is applied immediately afterward to suppress the mixer-induced image frequencies. Two preamplifiers, AD HMC499LC4 and AD HMC1131, are deployed to guarantee a sufficiently high power at the input of the active antenna array such that it can be pushed close toward saturation. The 64-element active array—the Anokiwave AWMF-0129—is then configured to transmit the signal OTA toward the horn antenna of the ORX. Both antennas are mounted on different electrical tripods capable of providing the required rotation in azimuth and elevation. For DPD learning and verification, both antennas are perfectly aligned when transmitting/receiving, as noted in Section II-C. For simplicity, the beamforming angle of the TX antenna array is set at 0°. The combined learning signal captured by the ORX is then attenuated and fed back to a Keysight UXR0402AP oscilloscope, which is acting as the digitizer. Finally, the digital signal is sent to a host PC for further processing and/or performance assessments. We note that the Anokiwave AWMF-0129 does not allow for actual hardware-based combiners for feedback, and hence, we adopt the carefully aligned OTA ORX to mimic such hardware-based processing.

The modulated signals adopted in the following experiments are 3GPP 5G NR Release-15 FR-2 compliant OFDM waveforms, with the subcarrier spacing (SCS) and RB allocation specified in each particular experiment. The sampling rate of the signals and DPD execution is 2 GHz. As an additional ingredient, which aims at pushing the performance boundaries of the DPD system, we also consider a wider, nonstandard-compliant, signal bandwidth of 800 MHz in some of the experiments. This is obtained by doubling the number of active subcarriers and also the OFDM waveform processing FFT size, compared to the standard-compliant 400 MHz signal. In addition, the initial PAPR of the generated signals is approximately 12 dB when measured at the 0.01% point of the instantaneous PAPR complementary cumulative distribution function (CCDF) [2]. This value is then limited through well-known iterative clipping and filtering-based processing to 8 dB, measured also at the 0.01% point of the instantaneous PAPR CCDF. Clipping the waveform, however, strictly speaking degaussianizes the input signal distribution, as it basically

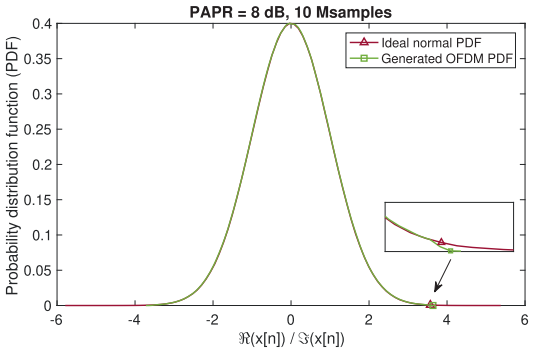


Fig. 3. PDF of the generated 5G NR FR-2 OFDM waveform with 400 MHz carrier bandwidth, including also PAPR mitigation to 8 dB through iterative clipping and filtering. The shown PDF reflects the real part of the complex I/Q waveform, while that of the imaginary part is essentially identical. For reference, also the theoretical normal PDF is shown.

removes the high- and low-end values of the amplitude PDF. The difference to the Gaussian distribution is nevertheless very small, as can be observed through the illustration in Fig. 3, which shows the PDF of the generated NR-compliant OFDM signal with clipping and filtering versus an ideal Gaussian PDF. Finally, after the PAPR mitigation, additional time-domain windowing is also applied to suppress the inherent OFDM signal sidelobes.

In each DPD iteration, a block of $K = 20\,000$ pseudorandomly generated OFDM signal samples is circularly transmitted, received, and utilized to estimate the DPD coefficients. The K-sized closed-loop error signal vector, $\mathbf{e}_k$, is generated as the subtraction between the transmitted and received data samples and essentially contains the prevailing PA distortion samples. This term is then used as an input of (5) or (6) to estimate the DPD coefficients. Such transmit/receive process is then repeated until the algorithms reach convergence. In all the experiments, the MP DPD model is configured with a ninth-order polynomial (i.e., $P = 9$) and with four taps of memory (i.e., $M = 4$). The DPD coefficients are initialized as $\boldsymbol{\alpha} = [1\,0\cdots 0]^T$ such that just the linear term is obtained after predistortion in the first iteration. In the measurements where the ICM is precalculated off-line, the CM is estimated from a 500 ksamples sequence and inverted before the DPD processing. It will remain fixed for the rest of the iterations unless otherwise mentioned.

B. Baseline DPD Performance

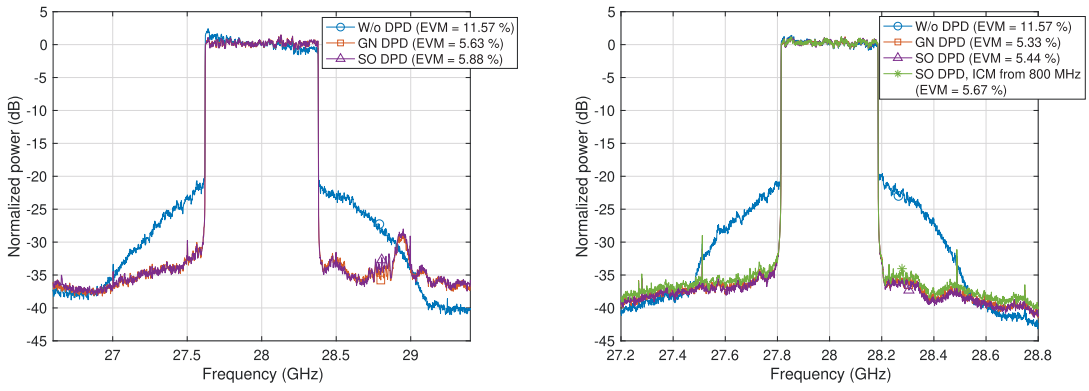
In this section, we study the linearization performance of the SO and the reference GN learning methods. The measurement results are carried out with both signal bandwidths of 800 MHz (120 kHz SCS, 528 RBs) and 400 MHz (120 kHz SCS, 264 RBs), measured at a highly nonlinear operation point of the active phased array. Specifically, this experiment is performed with effective isotropic radiated power (EIRP) of approximately +40.5 dBm, corresponding to an input power of approximately −6 dBm. This leads to an initial TRP-ACLR of approximately +22 dBc and an EVM of some 11.6%, reflecting a highly nonlinear operation point.

First, Fig. 4(a) presents the obtained linearization results with a signal bandwidth of 800 MHz for both SO and reference GN learning methods. We assume for the SO method a precalculated ICM, which is then kept fixed through the DPD iterations. The GN learning rule, in contrast, calculates the ICM estimate $(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k)^{-1}$ in every DPD iteration. As can be seen from the figure, the linearization performance of the SO method is practically matching that of the GN, with TRP-ACLR numbers of about +32.5 dBc and EVM values below 6%. This performance is also achieved with the SO solution regardless of the rough ICM off-line estimation. Secondly, Fig. 4(b) presents the results with a signal bandwidth of 400 MHz. Similar conclusions can be drawn regarding the behavior of the SO and GN learning methods. In this case, the TRP-ACLR is around +35 dBc and the EVM is below 5.5% for both SO and GN methods. What is more important, we also include a third measurement that shows the linearization behavior of the SO learning method when the ICM is calculated from the previous 800 MHz signal (i.e., the same as the one utilized in Fig. 4(a)). As shown in the figure, the DPD performance is only slightly degraded in comparison to the normal SO learning rule, with the TRP-ACLR and EVM being around +34 dBc and 5.7%, respectively. This is because the ICM has been obtained from a wider signal than the one being transmitted. However, the SO with an ICM obtained from the 400 MHz signal would not work to linearize the wider 800 MHz signal. We finally note that all measured methods satisfy the TRP-ACLR limit of +28 dBc and EVM limit of 8% (for 64-QAM), as stated in 3GPP specifications [7].

C. Reducing the ICM Dimensionality

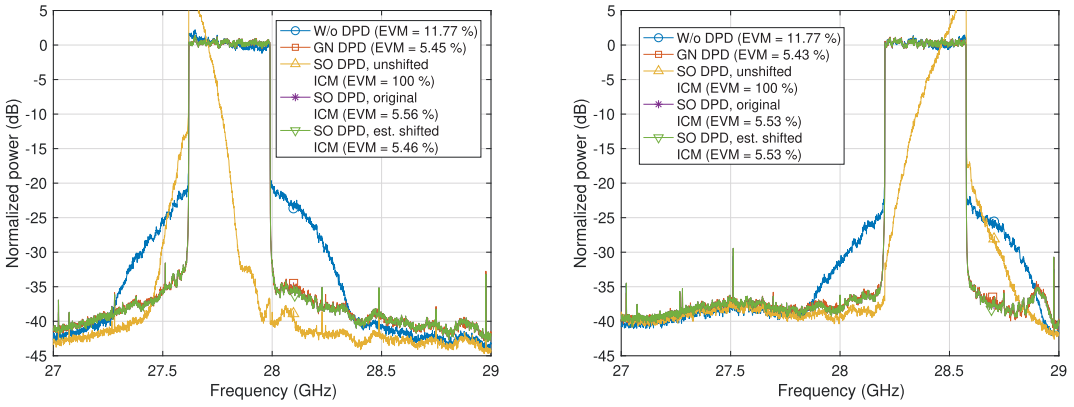
In this second experiment, we study and highlight the obtained numerical precision of applying the method, which removes rows/columns from the ICM to get a new covariance set for a DPD system with reduced parameterization. We recall that the rows/columns that are to be removed from the ICM are shifted to the last row/column, and then, this algorithm is repeated iteratively until all the targeted elements are successfully removed.

To this end, we perform the following experiment. We first generate an ICM of size 36×36 corresponding to $P = 11$ and $M = 6$. We then execute the proposed algorithm described in Section III-A iteratively until $l = 16$ rows and columns are removed, leading to a final ICM of size 20×20 elements—corresponding to $P = 9$ and $M = 4$. Parallel to that, we mathematically calculate, using (7), the corresponding ICM of size 20×20 (also corresponding to $P = 9$ and $M = 4$) and compare it to the estimated version. The obtained numerical error when comparing the two versions is in the order of 10^{-13}. This order of magnitude can be considered negligible in any DPD implementation, thus demonstrating the effectiveness of the considered technique. A similar experiment is next repeated for reducing the DPD parameterization down to $P = 7$ and $M = 3$, corresponding to $l = 24$. The numerical error in this case between the ideal and



(a) (b)

Fig. 4. Baseline OTA linearization results at EIRP $\approx +40.5$ dBm for the SO and GN learning methods with (a) 800 MHz bandwidth and (b) 5G-NR 400 MHz signal bandwidth, also showing the SO learning rule with the ICM estimated from the 800 MHz signal.



(a) (b)

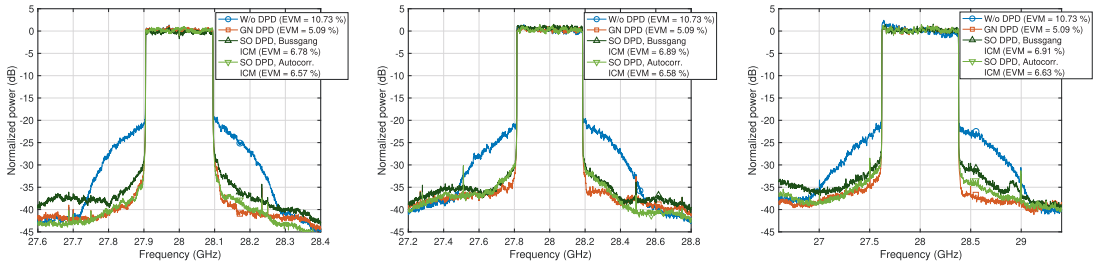
Fig. 5. 400 MHz 5G-NR OTA linearization performance at EIRP $\approx +40.5$ dBm for the SO method, with emphasis on the ICM shifting technique proposed in Section III-B, when a frequency shift of (a) -200 and (b) $+400$ MHz is applied. The linearization performance of the GN method is also included for reference. Note that the "SO DPD, original ICM" and "SO DPD, est. shifted ICM" curves practically overlap.

estimated ICMs is in the order of 10^{-11}, reflecting again very high accuracy. Finally, we note that no PSDs are visualized along this experiment, as with such very low numerical errors, the PSDs corresponding to the calculated and estimated ICMs are essentially completely overlapping.

D. Frequency Shifting the ICM

In this third experiment, we study the linearization performance of the SO learning rule in combination with the ICM shifting method proposed in Section III-B. The measurements are carried out with a signal bandwidth of 400 MHz (120-kHz SCS, 264 RBs), while the EIRP, the array's input power, and thus initial TRP-ACLR and initial EVM values are maintained as before.

The experiments to test the proposed method are carried out as follows. First, the input 5G NR 400 MHz baseband signal is generated, and then, two different example digital frequency shifts of -200 and $+400$ MHz are applied to it. This generates the resulting shifted RF signals, which are centered at 27.8 and 28.4 GHz. We report the linearization performance of the SO solution with: 1) the unshifted ICM; 2) the original ICM (i.e., estimated from the shifted signal); and 3) the ICM generated with the proposed method, presented in Section III-B. The performance of the GN solution is also presented as a reference. As can be seen from the obtained results in Fig. 5, the SO with the unshifted ICM does not converge to any solution since the ICM is not able to describe the statistics of the shifted signal. Second, the SO with the original ICM provides the optimal linearization performance, essentially equal to GN, since the ICM is directly calculated from the shifted signal. Third, the SO with the ICM estimated using the proposed method in Section III-B also provides optimal performance, despite the substantially lower associated complexity. The TRP-ACLR and the EVM



(a) (b) (c)

Fig. 6. OTA linearization performance at EIRP $\approx +40.5$ dBm for the SO method when the ICM is estimated through the two novel methods presented in Section IV, with the signal bandwidth of (a) 200, (b) 400, and (c) 800 MHz. The GN method is also measured and shown for reference.

numbers for the proposed method are in both cases maintained above +31 dBc and below 5.6%, respectively, thus satisfying the 3GPP standards [7] at FR-2. This performance is equal to that of the GN method, which involves heavy complexity in each DPD iteration. These results verify the effectiveness of the proposed ICM shifting method in terms of linearization performance.

E. Estimating ICM From the Autocorrelation Function

This fourth experiment validates the two novel methods presented in Section IV, which estimates the ICM from the autocorrelation function of the input data. These techniques are tested with three different signal bandwidths of 200 MHz (120 kHz SCS, 132 RBs), 400 MHz (120 kHz SCS, 264 RBs), and 800 MHz (120 kHz, 528 RBs). In all cases, the ICM is estimated using the proposed approach described in Section IV and then injected in the SO learning rule. The GN method is also measured for reference, while all the remaining system parameters are maintained as explained before.

The obtained measurement results are presented in Fig. 6. With a bandwidth of 200 MHz, the SO DPD with autocorrelation ICM estimation achieves an excellent linearization performance, very close to that of the GN approach. The SO DPD with the Bussgang ICM estimation lies somewhat behind but still obtains a good amount of linearization. This difference is reduced when considering the wider bandwidths of 400 and 800 MHz, in which the performance of the proposed novel solutions is very close to each other and also to the reference GN model. It is noted that there is no direct theory-based reason why the performance gap between the autocorrelation-based ICM estimation and the Bussgang ICM estimation methods is largest in the case of 200 MHz bandwidth. The difference can be stemming, e.g., from a slight change in the hardware setup, or by a minor movement in the position of the person handling the measurements. In all cases, the SO solution in combination with the methods proposed in Section IV successfully satisfies the +28 dBc TRP-ACLR and the 8% EVM limits while showing a remarkable complexity reduction in estimating the ICM.

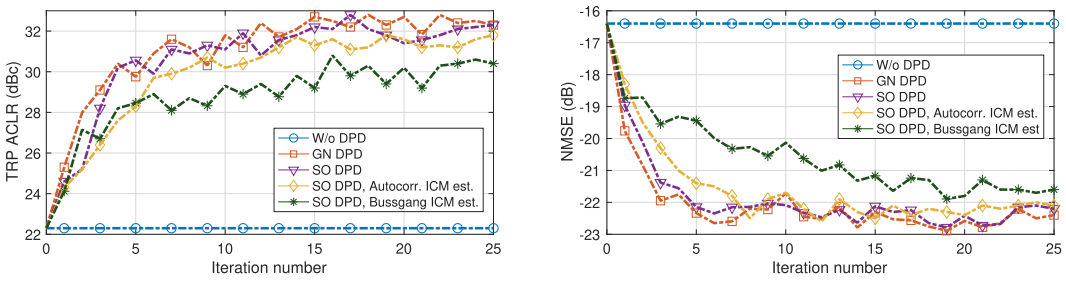
F. Convergence Analysis

We continue the experimental results with a convergence analysis of the proposed DPD methods in a static waveform scenario. For this analysis, we consider the signal bandwidth of 800 MHz (120-kHz SCS, 528 RBs, 625 ksamples) and present the convergence behavior of the SO method when: 1) ICM is estimated normally from the input signal; 2) ICM is estimated with the autocorrelation method presented in Section IV-A; and 3) ICM is estimated with the Bussgang-based method presented in Section IV-B. The convergence speeds of the remaining methods are the same as that of 1), and thus, they are not shown separately. The remaining DPD parameters are maintained as stated above.

The convergence behavior of the DPD models is presented in Fig. 7, in terms of the TRP-ACLR and the NMSE. As can be seen from the figure, the TRP-ACLR convergence of the GN, the SO with classical ICM calculation, and the SO with autocorrelation-based ICM estimation is very similar, reaching the steady-state performance in around 15 CL iterations, after which the DPD behavior stabilizes at around +32 dBc TRP-ACLR. The convergence speed of the SO with the Bussgang ICM estimation is slightly slower, reaching the steady state in around 20 block iterations. However, the final values of TRP-ACLR are similar compared to the previous methods. This behavior is expected since the Bussgang-based method provides only an approximate ICM. Similar conclusions can be drawn from the second figure, which shows that the NMSE obtained with the different methods. These results show the favorable performance of the proposed SO solutions, which together with the achieved complexity reductions demonstrates very appealing performance–complexity tradeoffs.

G. DPD Learning Complexity With Different ICM Methods

In this final subsection, we gather the quantitative DPD learning complexity numbers obtained with the proposed ICM estimation methods and corresponding to the DPD parameterization of $P = 9$, $M = 4$, $C = 20$, $l = 5$, and $K = 20\,000$. Here, we measure the complexity in terms of average number of real multiplications per linearized sample, assuming that the DPD learning algorithms are executed over 15 closed-loop iterations—which is generally the number of iterations that the



(a) (b)

Fig. 7. OTA convergence performance with 800 MHz waveform bandwidth and EIRP $\approx +40.5$ dBm for the SO method and its variants, in terms of (a) TRP-based ACLR and (b) NMSE. The GN method is also measured and shown for reference.

TABLE II
AVERAGE LEARNING PATH COMPLEXITIES OF THE PROPOSED METHODS PER LINEARIZED SAMPLE AND THE RELATIVE COMPLEXITY REDUCTIONS WITH RESPECT TO THE REFERENCE GN APPROACH

DPD method	Average learning comp. (real mul. per lin. sample)	Percentage red. wrt. GN
Reference GN	1,681	0 %
SO + classical ICM calc.	186.85	88.9 %
SO + ICM subset	80.10	95.2 %
SO + freq. shift ICM	80.08	95.2 %
SO + autocorr. est. ICM	81.54	95.1 %
SO + Bussgang's est. ICM	81.42	95.1 %

models need to converge, as demonstrated in the earlier subsections. The obtained quantitative complexity results, building on the expressions in Table I, together with the corresponding relative complexity reduction with respect to the reference GN method, are presented in Table II.

As can be seen from the table, excellent complexity savings can be obtained through the proposed solutions. In all cases, the numerical average complexity is drastically reduced when comparing with the reference GN method. This is also reflected in the achieved percentage reduction, which is in most cases larger than 90%. Keeping in mind that all the proposed methods satisfy the 3GPP transmit waveform quality specifications [7], as shown through the obtained results in Section VI-B–E, the combination of efficient linearization performance and low processing and learning complexity is paving the way toward fast and continuous DPD adaptation with on-chip real-time implementations in commercial systems.

VII. CONCLUSION

In this article, various methods to efficiently estimate the inverse of the input data CM were proposed and combined with SO closed-loop learning in a DPD-based linearization context. The inherent low complexity of the SO learning combined with the proposed methods, allowed for remarkably reduced complexities in the DPD system, while maintaining a similar linearization performance compared to state-of-the-art solutions. To substantiate this, complexity analyses dealing with the proposed solutions were performed, and thorough RF measurement results at 28 GHz mmW band, featuring a 64-element active antenna array, were presented. The obtained results, both in terms of performance and complexity, indicated very favorable performance–complexity tradeoffs of the proposed methods when comparing against the state of the art. The proposed methods are thus promising candidates for linearizing mmW phased-array transmitters as well as cellular terminal transmitters, the processing complexity being a key concern in both applications.

REFERENCES

[1] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.
[2] 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, 4th Quart., 2013.
[3] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
[4] S. Afsardoost, T. Eriksson, and C. Fager, "Digital predistortion using a vector-switched model," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 4, pp. 1166–1174, Apr. 2012.
[5] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, Nov. 2001.
[6] A. Zhu, "Digital predistortion and its combination with crest factor reduction," in *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*, F.-L. Luo, Ed. Cambridge, U.K.: Cambridge Univ. Press, 2011, ch. 9, pp. 244–279.
[7] *NR: Base Station (BS) Conformance Testing, Part 2*, document 3GPP TS 38.141-2, Version 15.1.0, Release 15, Mar. 2019.
[8] 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.
[9] C. Mollén, E. G. Larsson, U. Gustavsson, T. Eriksson, and R. W. Heath, Jr., "Out-of-band radiation from large antenna arrays," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 196–203, Apr. 2018.
[10] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, Dec. 2019.
[11] M. Abdelaziz, L. Anttila, A. Brihuega, F. Tufvesson, and M. Valkama, "Digital predistortion for hybrid MIMO transmitters," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 445–454, Jun. 2018.
[12] X. Liu *et al.*, "Beam-oriented digital predistortion for 5G massive MIMO hybrid beamforming transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 7, pp. 3419–3432, Jul. 2018.
[13] A. Brihuega *et al.*, "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, Sep. 2020.

[14] H. Holma, A. Toskala, and T. Nakamura, *5G Technology: 3GPP New Radio*, 1st ed. Hoboken, NJ, USA: Wiley, 2019.
[15] A. S. Tehrani, H. Cao, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.
[16] F. Mkadem, A. Islam, and S. Boumaiza, "Multi-band complexity-reduced generalized-memory-polynomial power-amplifier digital predistortion," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 6, pp. 1763–1774, Jun. 2016.
[17] A. Zhu and T. J. Brazil, "An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers," in *Proc. IEEE Annu. Wireless Microw. Technol. Conf.*, Dec. 2006, pp. 1–5.
[18] A. Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned Volterra series," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 12, pp. 563–565, Dec. 2004.
[19] H. Wang, G. Li, C. Zhou, W. Tao, F. Liu, and A. Zhu, "1-bit observation for direct-learning-based digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2465–2475, Jul. 2017.
[20] A. Zhu, M. Wren, and T. J. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers," in *IEEE MTT-S Int. Microw. Symp. Dig.*, vol. 2, Jun. 2003, pp. 787–790.
[21] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 1, pp. 280–293, Jan. 2018.
[22] N. Guan, N. Wu, and H. Wang, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 921–923, Oct. 2018.
[23] P. P. Campo, V. Lampu, A. Meirhaeghe, J. Boutellier, L. Anttila, and M. Valkama, "Digital predistortion for 5G small cell: GPU implementation and RF measurements," *J. Signal Process. Syst.*, vol. 92, pp. 475–486, May 2020.
[24] L. Ding, R. Raich, and G. T. Zhou, "A Hammerstein predistortion linearization design based on the indirect learning architecture," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, May 2002, pp. 2689–2692.
[25] P. P. Campo, L. Anttila, D. Korpi, and M. Valkama, "Cascaded spline-based models for complex nonlinear systems: Methods and applications," *IEEE Trans. Signal Process.*, vol. 69, pp. 370–384, 2020.
[26] X. Hong, Y. Gong, and S. Chen, "A Wiener model for memory high power amplifiers using B-spline function approximation," in *Proc. 17th Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2011, pp. 1–5.
[27] 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.
[28] P. P. Campo *et al.*, "Gradient-adaptive spline-interpolated LUT methods for low-complexity digital predistortion," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no, 1, pp. 336–349, Jan. 2021.
[29] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 3, pp. 980–987, Mar. 2017.
[30] Y. Ma, Y. Yamao, Y. Akaiwa, and C. Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 11, pp. 3961–3973, Nov. 2013.
[31] F.-L. Luo, *Digital Front-End in Wireless Communications and Broadcasting: Circuits and Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[32] L. Ding and G. T. Zhou, "Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 156–162, Jan. 2004.
[33] H. Enzinger, K. Freiberger, G. Kubin, and C. Vogel, "Baseband Volterra filters with even-order terms: Theoretical foundation and practical implications," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 220–224.
[34] N. Lashkarian, J. Shi, and M. Forbes, "A direct learning adaptive scheme for power-amplifier linearization based on Wirtinger calculus," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 12, pp. 3496–3505, Dec. 2014.
[35] 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.
[36] L. Guan and A. Zhu, "Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers," in *Proc. Workshop Integr. Nonlinear Microw. Millim.-Wave Circuits*, Vienna, Austria, Apr. 2011, pp. 1–4.
[37] P. Jardin and G. Baudoin, "Filter lookup table method for power amplifier linearization," *IEEE Trans. Veh. Technol.*, vol. 56, no. 3, pp. 1076–1087, May 2007.
[38] P. P. Campo *et al.*, "Closed-loop sign algorithms for low-complexity digital predistortion: Methods and performance," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 1, pp. 1048–1062, Jan. 2021.
[39] Q. Luo, X. Zhu, C. Yu, and W. Hong, "Single-receiver over-the-air digital predistortion for massive MIMO transmitters with antenna crosstalk," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 301–315, Jan. 2020.
[40] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G massive MIMO wireless transmitters based on indirect identification of power amplifier behavior with OTA tests," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 1, pp. 316–328, Jan. 2020.
[41] X. Liu, W. Chen, L. Chen, F. M. Ghannouchi, and Z. Feng, "Linearization for hybrid beamforming array utilizing embedded over-the-air diversity feedbacks," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 12, pp. 5235–5248, Dec. 2019.
[42] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU Press, 2012, pp. 238–239.
[43] I. Reed, "On a moment theorem for complex Gaussian processes," *IRE Trans. Inf. Theory*, vol. 8, no. 3, pp. 194–195, Apr. 1962.
[44] T. Araujo and R. Dinis, "On the accuracy of the Gaussian approximation for the evaluation of nonlinear effects in OFDM signals," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 346–351, Feb. 2012.
[45] J. Minkoff, "The role of AM-to-PM conversion in memoryless nonlinear systems," *IEEE Trans. Commun.*, vol. COM-33, no. 2, pp. 139–144, Feb. 1985.
[46] I. Cosme, I. Fernandes, J. Carvalho, and S. X.-D. Souza, "Block recursive matrix inverse," Dept. Engenharia Computaçao Automaçao, Univ. Federal Rio Grande Norte, Natal, Brazil, Tech. Rep., 2016.
[47] C. F. Van Loan, "The ubiquitous Kronecker product," *J. Comput. Appl. Math.*, vol. 123, nos. 1–2, pp. 85–100, 2000.
[48] W.-H. Steeb and Y. Hardy, *Matrix Calculus and Kronecker Product: A Practical Approach to Linear and Multilinear Algebra*, 2nd ed. Singapore: World Scientific, 2011.
[49] L. Guan and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 4, pp. 866–872, Apr. 2010.
[50] S. S. Haykin, *Adaptive Filter Theory*. London, U.K.: Pearson, 2008, pp. 226–227.

Pablo Pascual Campo (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunications and electrical engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 2016 and 2018, respectively. He is currently pursuing the D.Sc. degree at the Department of Electrical Engineering, Tampere University, Tampere, Finland.

His research interests include digital predistortion, full-duplex systems and applications, and signal processing for wireless communications at the mmWave bands.

Lauri Anttila received the D.Sc. (Tech.) degree (Hons.) from the Tampere University of Technology (TUT), Tampere, Finland, in 2011.

Since 2016, he has been a University Researcher with the Department of Electrical Engineering, Tampere University. From 2016 to 2017, he was a Visiting Research Fellow with Aalto University, Finland. His research interests are in radio communications and signal processing, with a focus on the radio implementation challenges in systems such as 5G, full-duplex radio, and large-scale antenna systems.

Vesa Lampu received the B.Sc. and M.Sc. degrees in electrical engineering from Tampere University, Tampere, Finland, in 2017 and 2019, respectively, where he is currently pursuing the D.Sc. degree.

He has been with the Department of Electrical Engineering, Tampere University, since 2018. His research focuses on energy-efficient multiple-input–multiple-output systems. He has also been involved in research on nonlinear system identification, with applications in full-duplex and digital predistortion techniques.

Markus Allén received the B.Sc., M.Sc., and D.Sc. degrees in communications engineering from Tampere University of Technology, Tampere, Finland, in 2008, 2010, and 2015, respectively.

He is currently with the Department of Electrical Engineering, Tampere University, as a University Instructor. His current research interests include software-defined radios, 5G-related RF measurements, and digital signal processing for radio transceiver linearization.

Yan Guo received the B.E. degree in information science and engineering from East China Jiaotong University, Nanchang, Jiangxi, China, in 2007, the M.E. degree in communication and information systems from Southeast University, Nanjing, China, in 2011, and the Ph.D. degree in electronic engineering from University College Dublin (UCD), Dublin, Ireland, in 2016.

He is currently a Principal Research Engineer and a Research Team Leader with the Wireless Terminal Chipset Algorithm Department, Hisilicon, Huawei. His main research interests include nonlinear behavioral modeling and digital linearization techniques for RF front-end of broadband UE.

Neng Wang received the Ph.D. degree in electrical engineering from Queen's University, Kingston, ON, Canada, in 2005.

In 2006, he was with the Communications Research Centre, Ottawa, ON, Canada, as a Research Scientist, worked on signal processing for wireless communications. From 2007 to 2009, he was with Nortel Networks, Richardson, TX, USA, as a Member of Technical Staff, worked on 3GPP LTE and 3GPP2 standardization. From 2010 to 2016, he was with Qualcomm Corp R&D, Beijing, China, as a Senior Staff Engineer/Manager, worked on LTE modem algorithms, 4G Evolution, and 5G PHY/MAC systems design and standardization. In 2016, he joined the Huawei Research Center, Lund, Sweden, as an Algorithm Expert/Lead, working on 5G modem algorithms. His research interests include general areas of wireless communications PHY/MAC systems and signal processing.

Mikko Valkama (Senior Member, IEEE) received the D.Sc. (Tech.) degree (Hons.) from the Tampere University of Technology (TUT), Tampere, Finland, in 2001.

In 2003, he was a Visiting Post-Doctoral Research Fellow with SDSU, San Diego, CA, USA. He is currently a Full Professor and the Department Head of electrical engineering with Tampere University (TAU), Tampere. His research interests include radio communications, radio localization, and radio-based sensing, with emphasis on 5G and beyond mobile radio networks.

PUBLICATION 5

Nonlinear digital cancellation in full-duplex devices using spline-based Hammerstein model

P. Pascual Campo, D. Korpi, L. Anttila and M. Valkama

*2018 IEEE Global Communications Conference (GLOBECOM)*2018, 1–7
DOI: 10.1109/GLOCOMW.2018.8644362

Publication reprinted with the permission of the copyright holders

Nonlinear Digital Cancellation in Full-Duplex Devices using Spline-Based Hammerstein Model

Pablo Pascual Campo, Dani Korpi, Lauri Anttila, and Mikko Valkama

Laboratory of Electronics and Communications Engineering, Tampere University of Technology, Finland
e-mail: pablo.pascualcampo@tut.fi, dani.korpi@tut.fi, lauri.anttila@tut.fi, mikko.e.valkama@tut.fi

***Abstract*—In this paper, a novel digital self-interference canceller based on a Hammerstein adaptive filter is proposed and examined. The proposed system consists of a spline-interpolated lookup table to model the nonlinear power amplifier, followed by a linear filter accounting for the impulse response of the linear self-interference channel. The gradient descent based parameter learning algorithms are derived, which estimate the spline control points and the filter coefficients in a decoupled manner. The proposed digital canceller leads to a complexity reduction of 77% when compared to the existing state-of-the-art solutions. Performance evaluations using measured data from a complete inband full-duplex prototype system operating at 2.4 GHz ISM band show the effectiveness of the proposed technique, demonstrating that it obtains similar cancellation performance as the existing state-of-the-art canceller, regardless of its lower complexity. The measured digital self-interference cancellation values are 45 dB, 43 dB and 38 dB with 20 MHz, 40 MHz and 80 MHz channel bandwidths, respectively. These results indicate that the complexity-accuracy trade-off of the proposed decoupled spline-based cancellation approach is very favorable. Owing to the resulting decrease in the computational complexity, the proposed digital cancellation technique brings inband full-duplex transceivers one step closer to commercial deployments.**

***Index Terms*—Spline interpolation, inband full-duplex, decoupled model, self interference, digital cancellation, nonlinear distortion, adaptive tracking**

I. Introduction

The practical feasibility of inband full-duplex communications, where individual radio devices transmit and receive simultaneously on the same frequency channel, has recently been proven by various research groups [1]–[5]. The main motivation for implementing such inband full-duplex systems stems from the resulting increase in spectral efficiency; the simultaneous transmission and reception facilitates a twofold increase in the data rate without requiring any additional bandwidth. Such an improvement is highly sought after especially in the heavily congested ultra high frequency (UHF) bands, and may turn out to be a key technology in the future wireless networks.

However, in order to make inband full-duplex technology commercially feasible, the problem of self-interference (SI) must be dealt with in an efficient manner [6], [7]. The SI is caused by the own transmit signal overlapping the desired received signal both temporally and spectrally due to the inband full-duplex operation, and the full-duplex device must be capable of suppressing it in order to operate properly. In principle, this can be done by subtracting the known transmit signal from the received signal, after accounting for the effects of the coupling channel. Since any full-duplex device must inevitably perform the SI cancellation procedure, it is crucial to explore ways of minimizing the involved computational complexity, while at the same time being able to provide the needed SI cancellation performance.

In this paper, we propose a novel digital SI cancellation solution, which utilizes an adaptive Hammerstein filter while modeling the nonlinearity with splines to significantly reduce the computational complexity compared to existing state-of-the-art solutions. In particular, we extend the real-valued spline adaptive filters from [8] to complex-valued form, and derive decoupled adaptive learning rules for the spline control points and the linear filter. To the best of our knowledge, this type of a spline-based digital SI canceller has not been proposed in the earlier literature.

Compared to the state-of-the-art solutions [2], [4], the proposed digital canceller achieves the same cancellation performance with 77% fewer computations. This is proven by evaluating the proposed canceller with measured SI signals from a complete inband full-duplex prototype system operating at 2.4 GHz ISM band and comparing its performance to that of the earlier solutions. The obtained results show that the spline-based Hammerstein digital canceller is a promising solution for bringing the inband full-duplex technology one step closer to a commercially viable implementation, with a very favorable cancellation performance–complexity trade-off.

The rest of this paper is organized as follows. In Section II, the proposed spline-based digital canceller is described, alongside with the parameter adaptation rules. Section III provides analysis of the computational complexity together with complexity comparison against a memory polynomial-based digital canceller, which represents the current state-of-the-art. After this, the performance of the digital cancellers is evaluated and compared in Section IV using measurement data. Finally, Section V concludes the paper.

Notation used in this paper

In this paper, vectors are represented with boldface lowercase letters. By default, all vectors consist of complex-valued elements represented as columns vectors (i.e., $\mathbf{x} \in \mathbb{C}^{N\times 1} = [x_0 \; x_1 \; \cdots \; x_{N-1}]^T$). Matrices are expressed with boldface capital letters (i.e., $\mathbf{A} \in \mathbb{C}^{N\times M}$). Ordinary transpose and Hermitian transpose of vectors are represented as $(\cdot)^T$ and $(\cdot)^H$, respectively. Moreover, the absolute value and floor operator are denoted by $|\cdot|$ and $\lfloor\cdot\rfloor$, respectively.

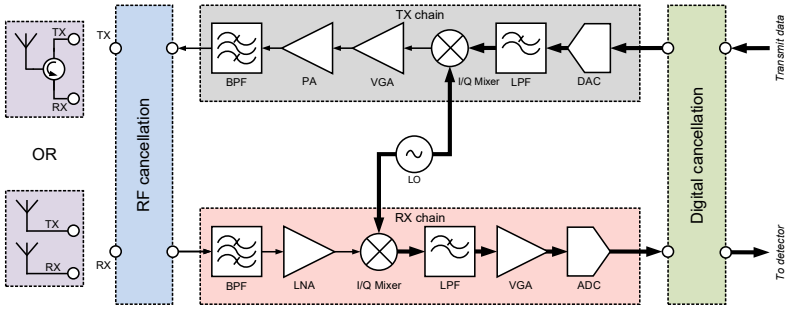


Fig. 1: Considered inband full-duplex device architecture.

II. System Model and Proposed Spline-based Canceller

A. Principles of Spline Interpolation

In spline interpolation, particular type of piece-wise polynomials under continuity and smoothness constraints are used to interpolate the input data [9]. With this technique, a nonlinear system can be modeled accurately even when using low-order piece-wise polynomials, as opposed to conventional polynomial models, where high orders are usually needed. The piece-wise polynomials are controlled through a set of control points being adapted according to the nonlinearity present in the system.

Spline curves can be generalized to any piece-wise degree, denoted here by P_{SP}. The spline curve is therefore a combination of $P_{\mathrm{SP}}+1$ spline segments between so-called knots, which define the boundaries of the individual splines. The individual region between the ith and $i+1$th knot is defined as the ith span. Moreover, each spline curve is characterized by a P_{SP}th-degree spline basis function $p_i^{P_{\mathrm{SP}}}(u_n)$, which is defined as [10]

$$p_i^{P_{\mathrm{SP}}}(u_n) = \frac{q_{i+P_{\mathrm{SP}}+1} - u_n}{q_{i+P_{\mathrm{SP}}+1} - q_{i+1}} p_{i+1}^{P_{\mathrm{SP}}-1}(u_n) + \frac{u_n - q_i}{q_{i+P_{\mathrm{SP}}} - q_i} p_i^{P_{\mathrm{SP}}-1}(u_n), \quad (1)$$

where q_i is the control point for the ith span. The abscissa value and the span are defined as

$$u_n = \frac{x[n]}{\Delta x} - \left\lfloor \frac{x[n]}{\Delta x} \right\rfloor, \quad (2)$$

$$i_n = \left\lfloor \frac{x[n]}{\Delta x} \right\rfloor + \frac{Q-1}{2}, \quad (3)$$

where $x[n]$ is the input signal of the spline system, Δx is the uniform distance between the consecutive knots, and Q is the total number of knots.

This function can be obtained recursively, starting from the spline polynomial of degree $P_{\mathrm{SP}} = 0$, defined as

$$p_i^0(u) = \begin{cases} 1 & \text{if } q_i \le u < q_{i+1}, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

For a more detailed explanation, refer, for instance, to [9].

As shown in [10], the output of a spline nonlinearity can be expressed with matrix notation as

$$s_i(u_n) = \mathbf{S}_n \mathbf{q}_n, \quad (5)$$

where $\mathbf{q}_n \in \mathbb{R}^{Q\times 1} = [q_0 \; q_1 \; \cdots \; q_{Q-1}]^T$ is a real-valued vector containing the control points and $\mathbf{S}_n \in \mathbb{R}^{1\times Q}$ is defined as

$$\mathbf{S}_n = [0 \; \cdots \; 0 \; \mathbf{u}_n^T\mathbf{C} \; 0 \; \cdots \; 0]. \quad (6)$$

Here, $\mathbf{u}_n^T\mathbf{C}$ represents the matrix multiplication between the abscissa vector $\mathbf{u}_n \in \mathbb{R}^{(P_{\mathrm{SP}}+1)\times 1} = [u_n^{P_{\mathrm{SP}}} \; u_n^{P_{\mathrm{SP}}-1} \; \cdots \; 1]^T$ and the spline basis matrix $\mathbf{C} \in \mathbb{R}^{(P_{\mathrm{SP}}+1)\times(P_{\mathrm{SP}}+1)}$, which is pre-computed and fixed. The resulting $(P_{\mathrm{SP}}+1)\times 1$ vector is then sorted in $\mathbf{S}_n$ starting from the span index i in the current iteration. This way, $\mathbf{u}_n^T\mathbf{C}$ is multiplied with the correct samples of $\mathbf{q}_n$ in the ith span. The spline basis matrix $\mathbf{C}$ is dependent on the type of splines used, as well as on the spline polynomial order. For example, for B-splines and $P_{\mathrm{SP}} = 2$, the spline basis matrix becomes [10]

$$\mathbf{C} = \frac{1}{2}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 1 & 0 \end{bmatrix}. \quad (7)$$

B. Proposed Spline-Based Hammerstein Canceller

The objective of the proposed canceller is to accurately reconstruct nonlinearly distorted SI within a full-duplex device whose block diagram is depicted in Fig. 1. Previous works in [2], [7], [11] have shown that Hammerstein-type models are, in general, accurate in reconstructing nonlinearly distorted SI but they involve substantial processing complexity. To this end, we propose using a decoupled Hammerstein model, which consists of a static nonlinearity representing the nonlinear power amplifier (PA), followed by a linear filter which models the linear SI channel. This linear SI channel includes the PA memory as well as the the overall linear coupling response from the PA output to the receiver digital baseband.

The proposed spline-interpolated lookup table for modeling and cancelling the SI is illustrated in Fig. 2, where the signal $x[n]$ corresponds to the input signal of the proposed structure,

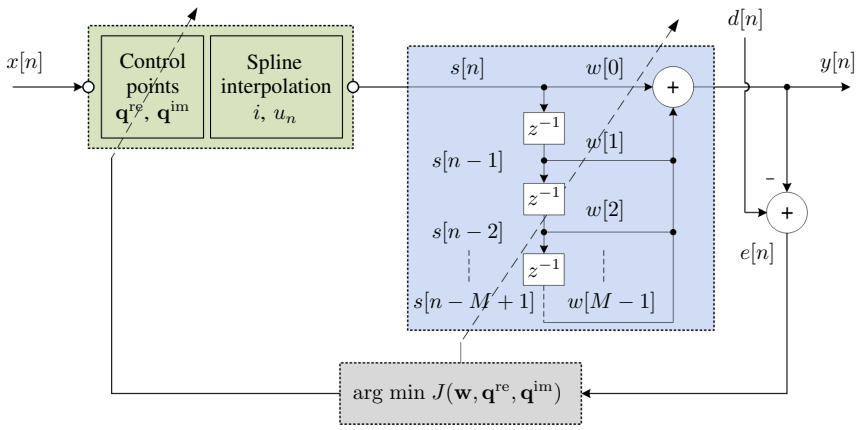


Fig. 2: Architecture of the proposed nonlinear decoupled spline-based Hammerstein self-interference canceller. Note that the pre-cursor taps are omitted from the diagram for clarity.

in this case the original transmit signal. However, this signal is complex-valued, whereas regular splines are real-valued. Since here the splines are used to model the nonlinear response of the PA, which only depends on the magnitude of the input signal, the real-valued splines can be used to model it by defining separate splines for the I and Q components and using $|x[n]|$ as the input of both splines. Consequently, the abscissa and span of both splines is now obtained by replacing $x(n)$ with $|x(n)|$ in (2) and (3), and omitting the term $\frac{Q-1}{2}$ from (3) as the abscissa is now non-negative. The splines modeling the real and imaginary (the I and Q) responses of the PA are therefore expressed as

$$s_1(u_n) = \mathbf{S_n}(1 + \mathbf{q}_n^{\mathrm{re}}), \tag{8}$$

$$s_2(u_n) = \mathbf{S}_n \mathbf{q}_n^{\mathrm{im}}, \tag{9}$$

where $\mathbf{q}_n^{\mathrm{re}}$ and $\mathbf{q}_n^{\mathrm{im}}$ contain the control points for the nth iteration.

By using the expression in (8) and (9), the complex-valued output of the nonlinear system can be written as

$$s[n] = x[n]\mathbf{S}_n(1 + \mathbf{q}_n^{\mathrm{re}} + j\mathbf{q}_n^{\mathrm{im}}) = x[n]\mathbf{S}_n(1 + \mathbf{q}_n), \tag{10}$$

where $\mathbf{q}_n$ is the overall complex-valued control point vector for the nth iteration. The signal $s[n]$ retains the phase information of the original signal via the multiplication by $x[n]$, while the two splines determine the real and imaginary responses of the nonlinear PA.

Compared to (5), we now define the control point vector $\mathbf{q}_n$ as a deviation from 1. This has several benefits, especially in the current application. First, if $\mathbf{q}_n$ is initialized as a zero vector, the spline will be linear (i.e., $s[n] = x[n]$) in the beginning of the algorithm learning period. This will effectively remove the gain ambiguity between the two blocks (linear filter and splines), and the linear filter will immediately start to converge to a state where it handles the gain. A second benefit is that the dynamic range of $\mathbf{q}_n$ is reduced, and thus a smaller number of bits is required in a fixed-point implementation.

In addition to the nonlinear distortion, also the memory effects need to be modeled. This includes the memory of the PA itself as well as the linear response of the rest of the overall SI channel, from PA output all the way to the receiver baseband. Considering this, the final output signal of the model is

$$y[n] = \mathbf{w}_n^H \mathbf{s}_n, \tag{11}$$

where $\mathbf{w}_n$ is the impulse response of the linear filter, defined as $\mathbf{w}_n \in \mathbb{C}^{M\times 1} = [w_n[0]\ w_n[1]\ \cdots\ w_n[M-1]]^T$, $\mathbf{s}_n \in \mathbb{C}^{M\times 1} = [s[n+M_{\mathrm{pre}}]\ \cdots\ s[n]\ \cdots\ s[n-M_{\mathrm{post}}]]^T$, M_{pre} is the number of pre-cursor taps, M_{post} is the number of post-cursor taps, and $M = M_{\mathrm{pre}} + M_{\mathrm{post}} + 1$.

C. Derivation of the Learning Rules

In order to estimate the unknown parameters of the system, namely the linear filter $\mathbf{w}_n$, and the control points $\mathbf{q}_n$, let us consider the error signal

$$e[n] = d[n] - y[n], \tag{12}$$

where $d[n]$ denotes the observed SI signal at receiver digital baseband, which the digital canceller aims at suppressing. At this point, the problem lies in estimating the values of the linear filter and the spline control points to minimize the error signal $e[n]$. This can be done by using the basic gradient descent solution, where the quantities are learned by adjusting them to the negative direction of the gradient of the cost function. Furthermore, to obtain a simple learning rule, only the instantaneous gradient is used, similar to the classical least mean squares (LMS) filter.

For the estimation of the memory model, the cost function is defined as the instantaneous squared error, expressed as

$$J(\mathbf{w}_n, \mathbf{q}_n) = e[n]e^*[n]. \tag{13}$$

It can then be noted that the memory is after the nonlinearity, and therefore the splines are not dependent on $\mathbf{w}_n$. As a result, the system with respect to $\mathbf{w}_n$ is simply a linear filter, where the traditional LMS learning rule can be used. Consequently, the learning rule for the memory filter becomes

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_\mathrm{w} \frac{\partial J(\mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{w}_n} = \mathbf{w}_n + \mu_\mathrm{w} e^*[n]\mathbf{s}_n, \tag{14}$$

where μ_w is the learning step-size.

The complex-valued control-points can be learned using a similar rule, which can be expressed as

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_\mathrm{q} \frac{\partial J(\mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{q}_n}, \tag{15}$$

where μ_q is the corresponding step-size. However, now the partial derivative is not previously known, and must be calculated in order to express the learning rule. Firstly, we can rewrite it as [12]:

$$\begin{aligned}
\frac{\partial J(\mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{q}_n} &= e^*[n]\frac{\partial e[n]}{\partial \mathbf{q}_n} + e[n]\frac{\partial e^*[n]}{\partial \mathbf{q}_n} \\
&= -e^*[n]\frac{\partial y[n]}{\partial \mathbf{q}_n} - e[n]\frac{\partial y^*[n]}{\partial \mathbf{q}_n} \\
&= -e^*[n]\left[\frac{\partial y[n]}{\partial \mathbf{q}_n^\mathrm{re}} + j\frac{\partial y[n]}{\partial \mathbf{q}_n^\mathrm{im}}\right] \\
&\quad - e[n]\left[\left(\frac{\partial y[n]}{\partial \mathbf{q}_n^\mathrm{re}}\right)^* + j\left(\frac{\partial y[n]}{\partial \mathbf{q}_n^\mathrm{im}}\right)^*\right].
\end{aligned} \tag{16}$$

Therefore, it suffices to determine the partial derivative of $y(n)$ with respect to the control points of the real and imaginary splines. Invoking elementary differentiation rules, they can be written as

$$\frac{\partial y[n]}{\partial \mathbf{q}_n^\mathrm{re}} = \mathbf{\Sigma}_n \mathbf{X}_n \mathbf{w}_n^*, \tag{17}$$

$$\frac{\partial y[n]}{\partial \mathbf{q}_n^\mathrm{im}} = j\mathbf{\Sigma}_n \mathbf{X}_n \mathbf{w}_n^*, \tag{18}$$

where $\mathbf{\Sigma}_n \in \mathbb{R}^{Q\times M} = \left[\mathbf{S}_{n+M_\mathrm{pre}}^T \ \cdots \ \mathbf{S}_n^T \ \cdots \ \mathbf{S}_{n-M_\mathrm{post}}^T\right]$, and $\mathbf{X}_n = \mathrm{diag}\left\{x[n+M_\mathrm{pre}], \ \ldots, \ x[n-M_\mathrm{post}]\right\}$.

Substituting these into (16), the partial derivative becomes

$$\frac{\partial J(\mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{q}_n} = -2e(n)\mathbf{\Sigma}_n \mathbf{X}_n^* \mathbf{w}_n. \tag{19}$$

Therefore, the learning rule of the control points can finally be written as

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_\mathrm{q} e(n)\mathbf{\Sigma}_n \mathbf{X}_n^* \mathbf{w}_n. \tag{20}$$

During the actual operation of the proposed canceller, both the linear filter and the control points are then estimated and updated in parallel as per (14) and (20).

The downside of the update in (20) is the high complexity in computing $\mathbf{\Sigma}_n \mathbf{X}_n^* \mathbf{w}_n$, especially if the filter $\mathbf{w}_n$ is long. However, the larger tap values of $\mathbf{w}_n$ are clearly the most significant ones in the update, and these tap values are usually those around the index $M_\mathrm{pre} + 1$. Thus, as a simplifying approximation, we propose to reduce the time span of the matrices, keeping only the entries corresponding to the most significant taps in $\mathbf{w}_n$. This way, the computation of $\mathbf{\Sigma}_n \mathbf{X}_n^* \mathbf{w}_n$ can be greatly simplified.

III. Computational Complexity Analysis

A. Complexity of Proposed Canceller

Let us next briefly consider the computational complexity of the proposed decoupled spline-based Hammerstein algorithm. For this purpose, the number of arithmetical operations required to run a single iteration of the algorithm is calculated, in a similar manner as done in [13]. However, for brevity, herein we only consider the number of multiplications since the computational cost of additions is negligible in comparison.

The complexity of the algorithm can be detailed based on the cancellation processing steps performed in each iteration.

- At the beginning of each iteration, the span index and the abscissa value (i_n and u_n) are computed. For this, it is necessary to calculate the absolute value of a complex number, whose analytical expression contains a square root. To avoid such a costly operation, the absolute value can be approximated as in [14]:

$$\begin{aligned}
|x[n]| = \alpha \ &\max\left\{|\mathrm{Re}\{x[n]\}|, |\mathrm{Im}\{x[n]\}|\right\} \\
+ \beta \ &\min\left\{|\mathrm{Re}\{x[n]\}|, |\mathrm{Im}\{x[n]\}|\right\},
\end{aligned} \tag{21}$$

  where α and β can be chosen based on selected approximation criterion, such as minimum RMS error for zero-mean signals.
- Next, the signals $s[n]$ and $y[n]$ must be calculated, as per (10) and (11).

Then, in order to update the coefficients of the learning rates, the following steps are followed:

- The linear filter $\mathbf{w}_{n+1}$ is updated as shown in (14).
- Finally, the control point vector $\mathbf{q}_{n+1}$ is updated according to (20). In the analysis we consider a time span of $\tau = 5$ taps for $\mathbf{w}_n$. This will considerably reduce the computational complexity of the vector update with an acceptable loss of cancellation.

The corresponding numbers of required multiplications per one iteration are collected in Table I. Note that in some cases it might not be necessary to update all the coefficients in each iteration, which obviously reduces the overall computational cost. In particular, as discussed below, it is often unnecessary to continuously update the spline control points as the nonlinear characteristics of the PA remain relatively constant over time.

B. Complexity of Memory Polynomial Canceller

The widely used memory polynomial (MP) model constitutes the current state of the art when modeling the behaviour of PAs in the context of inband full-duplex devices [2], [7], [11]. For this reason, it has been adopted as the reference benchmark for the proposed novel solution. Below, we shortly

TABLE I: Number of required arithmetic operations in each iteration of the spline-based digital cancellation algorithm.

Computation		Real multiplications
Cancellation	$s[n], y[n]$	$(P_{\mathrm{SP}}+1)^2+P_{\mathrm{SP}}+2Q+3M+4$
Coefficient updates	$\mathbf{w}_{n+1}$	$3M+2$
	$\mathbf{q}_{n+1}$	$Q(2\mathcal{T}+3)+3\mathcal{T}+2$
Total		$P_{\mathrm{SP}}(P_{\mathrm{SP}}+3)+\mathcal{T}(2Q+3)+5Q+6M+9$

review the principal processing structure of such MP-based digital SI canceller and address its computational complexity for reference.

The discrete-time MP model can be expressed as

$$y[n]=\sum_{\substack{p=1\\ p\ \mathrm{odd}}}^{P_{\mathrm{MP}}}\sum_{m=-M_{\mathrm{pre}}}^{M_{\mathrm{post}}} h_p[m]\psi_p[x[n-m]], \qquad (22)$$

where P_{MP} is the nonlinearity order of the MP model, $h_p[m]$ represents the overall effective coupling channel from the pth basis function perspective, and $\psi_p[x[n]]=|x[n]|^{p-1}x[n]$ is the pth-order basis function. In addition, in this model these basis functions need to be prewhitened or orthogonalized to ensure efficient learning [7]. After this step, the parameters of the MP model can be learned with the basic LMS algorithm, the use of which facilitates a fair comparison with the proposed solution. Such an algorithm is described in [11], wherein further details can be found. The total number of arithmetic operations can easily be calculated as $3(\frac{P_{\mathrm{MP}}+1}{2})^2+(3M+1)(P_{\mathrm{MP}}+1)+2$ real multiplications per one iteration, of which $3(\frac{P_{\mathrm{MP}}+1}{2})^2+(3M+2)(\frac{P_{\mathrm{MP}}+1}{2})$ are used for the actual cancellation processing while the coefficient updates require $3M(\frac{P_{\mathrm{MP}}+1}{2})+2$ real multiplications.

C. Comparison

For the purpose of comparing the aforementioned models, typical parameter values are chosen in accordance with the characteristics of each algorithm. When considering the decoupled spline-based Hammerstein model, the degree of the polynomial is chosen as $P_{\mathrm{SP}}=2$, which is sufficient to reach a SI cancellation performance similar to that of the MP model (see Section IV). The total number of knots is chosen as $Q=11$, and the time span for the matrix simplification is $\tau=5$. The overall memory length of the linear filter is chosen for both models as $M=M_{\mathrm{pre}}+M_{\mathrm{post}}+1=25+50+1=76$. Consequently, using Table I, the amount of computations per iteration can be calculated as 655 real multiplications for the proposed decoupled Hammerstein algorithm. Of this, the cancellation processing accounts for only 265 real multiplications.

As for the MP-based canceller, $P_{\mathrm{MP}}=11$ is chosen as the order of the nonlinearity, similar to [11]. As a result, the number of real multiplications to be performed within an iteration is 2858, according to the expression provided in the previous subsection. This demonstrates the large complexity reduction when using the novel decoupled spline-based Hammerstein approach, as nearly 77% less multiplications per iteration suffices to obtain a comparable cancellation performance. Moreover, when considering only the actual cancellation processing, the number of real multiplications per iteration is 1488 for the MP canceller. Therefore, under static conditions where it is not necessary to update any of the coefficients, the spline-based canceller can operate with 82% fewer multiplications.

However, in most practical cases, while the nonlinear behaviour of the PA remains more or less static, the wireless coupling channel is time-variant and the corresponding coefficients must be adapted continuously. This unveils another benefit of the proposed decoupled spline-based solution, as it considers the PA nonlinearity and the linear channel as separate entities, meaning that the channel coefficients can be updated while the spline control points are kept static after the initial learning phase. Taking this into account, it suffices for the proposed decoupled spline-based model to update only the memory filter during the actual operation, reducing the number of real multiplications per iteration to 495 with the example parameters. Recognizing that it is not possible to separate the effects of the linear coupling channel and the PA nonlinearity in the MP model, under this assumption the proposed spline-based canceller can operate with 82% fewer multiplications.

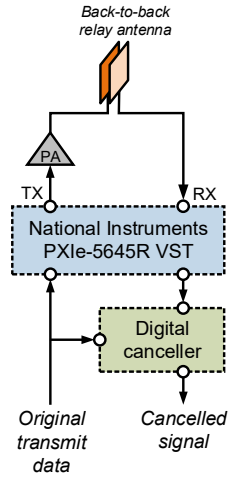
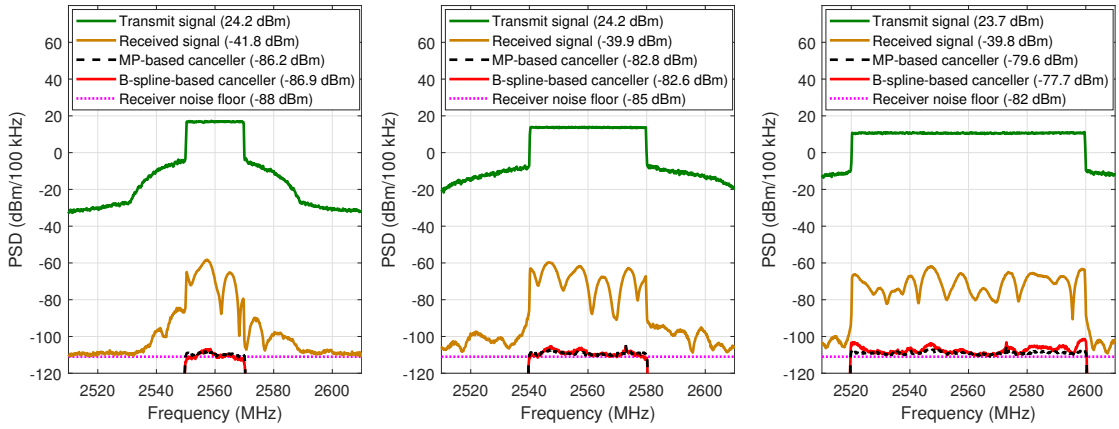


Fig. 3: Illustration of the overall RF measurement setup used in obtaining the evaluation data.

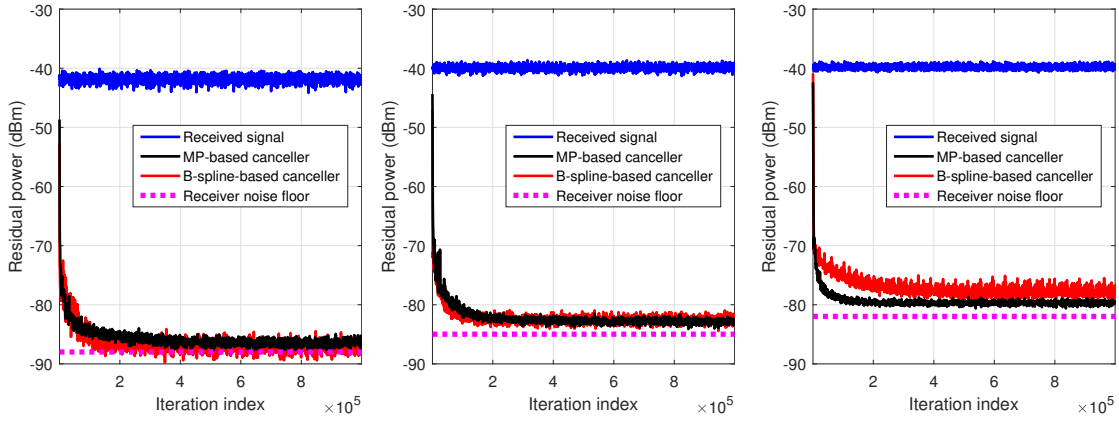
IV. Experimental results

In order to evaluate the cancellation performance of the proposed spline-based digital canceller, a similar RF measurement environment as in [4], reflecting a complete inband full-duplex prototype system, is used. The measured data consists of the observed SI signal obtained using the system depicted in Fig. 3, where a back-to-back relay antenna is used to provide the transmitter-receiver isolation in the analog domain. That



(a) (b) (c)

Fig. 4: PSDs of the overall signal after the different digital cancellers for (a) 20 MHz, (b) 40 MHz, and (c) 80 MHz instantaneous bandwidths.



(a) (b) (c)

Fig. 5: Residual powers with respect to the iteration index after the different digital cancellers for (a) 20 MHz, (b) 40 MHz, and (c) 80 MHz instantaneous bandwidths.

is, the RF canceller is omitted altogether, as opposed to the more generic full-duplex transceiver architecture depicted in Fig. 1. Since this results in less SI suppression before the analog-to-digital conversion, a significant amount of digital cancellation is required, making this data ideal for evaluating the performance of the proposed digital canceller. The more detailed measurement parameters are listed in Table II.

First, the power spectral densities (PSDs) of the SI signal after the different stages of the full-duplex device are shown in Fig. 4. With instantaneous bandwidths of 20 MHz and 40 MHz, the spline-based canceller obtains essentially the same amount of cancellation as the MP-based canceller, despite the substantial complexity reduction. In these two cases, the amount of digital cancellation with the proposed canceller is 45 dB and 43 dB, respectively, which indicates very high modeling accuracy. Keeping in mind that it uses 77% fewer multiplications than the MP model, the complexity-accuracy trade-off of spline-based modeling is very intriguing. With the widest considered bandwidth of 80 MHz, the MP-based canceller outperforms the proposed digital canceller by some 2 dB. Therefore, further work is still needed to ensure sufficient modeling accuracy of splines under very wideband operation. It should be noted, however, that the proposed canceller can suppress the SI by 38 dB even over 80 MHz, which is in most cases sufficient cancellation performance.

To evaluate the convergence behavior of the proposed digital canceller, Fig. 5 shows the residual power after the two digital cancellers, using the same data as in Fig. 4. For the

TABLE II: The essential RF measurement parameters.

Parameter		Value
Center frequency		2.56 GHz
Bandwidth		20–80 MHz
Transmit waveform		OFDM
Transmit power		24 dBm
RX losses		4 dB
Parameter estimation sample size		1 000 000
MP-based canceller	M_{pre}	25
	M_{post}	50
	P_{MP}	11
Spline-based canceller	M_{pre}	25
	M_{post}	50
	P_{SP}	2
	Q	11

20-MHz and 40-MHz SI signals, it can be observed that the proposed spline-based digital canceller converges faster than the MP-based canceller, thereby reaching the steady-state sooner. However, in accordance with the observations made in Fig. 4, the 80-MHz case seems to be somewhat more challenging for the spline-based digital canceller also in terms of convergence behaviour. Figure 5c shows that the spline-based canceller converges somewhat slower than the MP-based canceller, while also suffering from slightly higher steady-state error. Nevertheless, these results prove that with bandwidths of 40 MHz or less, the proposed canceller can retain the highest possible cancellation performance while using significantly less computational resources.

V. Conclusion

In this paper we proposed a novel digital self-interference canceller for inband full-duplex devices that relies on splines and decoupled modeling when reconstructing nonlinearly distorted self-interference signals at digital baseband. To the best of our knowledge, this is the first time splines have been applied to self-interference cancellation. The benefit of the proposed approach is a significant reduction in the number of computations required within the digital canceller while still retaining high cancellation performance. In particular, it was shown that the proposed solution can obtain cancellation performance comparable to the existing state-of-the-art digital canceller while using 77% fewer multiplications per iteration. The measured digital self-interference cancellation numbers, obtained in a complete inband full-duplex radio prototype system operating at 2.4 GHz ISM band, are 45 dB, 43 dB and 38 dB with 20 MHz, 40 MHz and 80 MHz channel bandwidths, respectively. Therefore, the spline-based digital canceller is an important step towards commercially feasible inband full-duplex devices, where processing complexity and energy consumption are of fundamental importance.

Acknowledgment

This work was supported by the Academy of Finland (under the projects #304147 "In-Band Full-Duplex Radio Technology: Realizing Next Generation Wireless Transmission", and #301820 Competitive Funding to Strengthen University Research Profiles), the Finnish Funding Agency for Innovation (Tekes, under the projects "5G Transceivers for Base Stations and Mobile Devices (5G TRx)" and "TAKE-5"), Nokia Networks, RF360 Europe, Pulse, Sasken, and Huawei Technologies, Finland.

References

[1] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[2] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *Proc. SIGCOMM'13*, Aug. 2013, pp. 375–386.

[3] D. Korpi, J. Tamminen, M. Turunen, T. Huusari, Y.-S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Full-duplex mobile device: Pushing the limits," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 80–87, Sep. 2016.

[4] D. Korpi, M. Heino, C. Icheln, K. Haneda, and M. Valkama, "Compact inband full-duplex relays with beyond 100 dB self-interference suppression: Enabling techniques and field measurements," *IEEE Transactions on Antennas and Propagation*, vol. 65, pp. 960–965, Feb. 2017.

[5] M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C. b. Chae, "Prototyping real-time full duplex radios," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 56–63, Sep. 2015.

[6] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[7] D. Korpi, "Full-duplex wireless: Self-interference modeling, digital cancellation, and system studies," Ph.D. dissertation, Tampere University of Technology, Dec. 2017.

[8] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties," *Signal Processing*, vol. 100, pp. 112–123, 2014.

[9] C. De Boor, *A practical guide to splines*. Springer-Verlag New York, 1978.

[10] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Nonlinear spline adaptive filtering," *Signal Processing*, vol. 93, no. 4, pp. 772–783, 2013.

[11] D. Korpi, Y.-S. Choi, T. Huusari, S. Anttila, L. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile inband full-duplex radio: algorithms and RF measurements," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015.

[12] D. P. Mandic and V. Su Lee Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.

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

[14] M. Frerking, *Digital Signal Processing in Communications Systems*. Springer, 1994.

PUBLICATION 6

Closed-loop sign algorithms for low-complexity digital predistortion

P. Pascual Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allén and M. Valkama

*2020 IEEE/MTT-S International Microwave Symposium (IMS)*2020, 841–844
DOI: 10.1109/IMS30576.2020.9223904

Publication reprinted with the permission of the copyright holders

Closed-Loop Sign Algorithms for Low-Complexity Digital Predistortion

Pablo Pascual Campo, Vesa Lampu, Lauri Anttila, Alberto Brihuega, Markus Allén, and Mikko Valkama
Department of Electrical Engineering, Tampere University, Finland

***Abstract*— In this paper, we study digital predistortion (DPD) based linearization with specific focus on millimeter wave (mmW) active antenna arrays. Due to the very large channel bandwidths and beam-dependence of nonlinear distortion in such systems, we propose a closed-loop DPD learning architecture, look-up table (LUT) based memory DPD models, and low-complexity sign-based estimation algorithms, such that even continuous DPD learning could be technically feasible. To this end, three different learning algorithms – sign, signed regressor, and sign-sign – are formulated for the LUT-based DPD models, such that the potential rank deficiencies, experienced in earlier methods, are avoided. Then, extensive RF measurements utilizing a state-of-the-art mmW active antenna array system at 28 GHz are carried out and reported to validate the methods. Additionally, the processing and learning complexities of the considered methods are analyzed, which together with the measured linearization performance figures allow to assess the complexity-performance tradeoffs. Overall, the results show that efficient mmW array linearization can be obtained through the proposed methods.**

***Keywords*— Array transmitters, nonlinear distortion, digital predistortion, mmW frequencies, sign algorithm, signed regressor, Hadamard, lookup table, ACLR.**

I. Introduction

The adoption of modern, spectrally efficient waveforms with high peak-to-average power ratio (PAPR), most notably OFDM, complicates operating power amplifiers (PAs) close to saturation [1]. To ensure a good power efficiency, while still emitting sufficiently low distortion levels, digital predistortion (DPD) based linearization is a well-known and widely-applied approach. In the existing literature, different DPD architectures and modeling methods have been widely studied, reflecting different processing complexities and linearization performance, with good overviews being available in [1]–[3].

In modern DPD use cases, particularly the active antenna array based base-stations of the emerging 5G New Radio (NR) networks at mmW bands (referred to as the frequency range 2, FR2) [4], the effective nonlinear distortion has been observed to be beam-dependent [5], and thus fast DPD adaptation is required. This issue, together with the very wide channel bandwidths [4], and thus DPD processing rates, calls for low-complexity DPD systems and parameter learning algorithms. Such methods are currently under intensive research and form also the topic of this paper.

To this end, the authors in [5] present a comprehensive overview of 5G NR array linearization, while [6] investigates a reduced complexity approach by utilizing the combined PA output signals together with a computationally efficient closed-loop (CL) learning rule to minimize the distortion in the main beam direction. In [7], in a more traditional single-antenna DPD context, the use of 1-bit observations in CL learning is considered, together with a sign-based Gauss-Newton (GN) learning algorithm. In [8], a GN signed regressor algorithm (SRA) is developed for real-valued feedback signals. The signed regressor matrix is, however, rank deficient, and thus an additional Walsh-Hadamard transformation is applied to make it invertible. In [9], an LUT based memory polynomial (MP) DPD with a sample-adaptive least mean squares (LMS) SRA is proposed. However, in this work each LUT in the MP structure is updated independently, making the solution sub-optimal. In [10], direct least squares (LS) and GN adaptations for linearly interpolated LUT-based Volterra models are proposed in indirect learning architecture (ILA) and CL context, respectively.

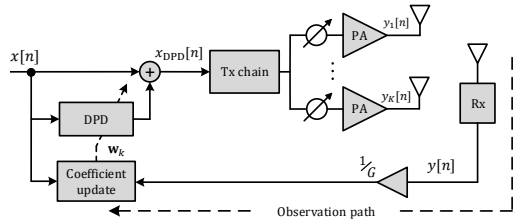


Fig. 1. The Closed-loop DPD system based on injection. G is the estimate of the complex linear gain of the chain and K is the number of antennas.

In this paper, contrary to the earlier CL works in [7]–[10], we adopt the so-called injection-based DPD structure [6], [11], and formulate various signed learning methods based on both GN and block-LMS, while adopting LUT-based memory DPD models. LUT-based structures are generally simpler than polynomial-type ones used in the reference works [6]–[8], allowing large reductions in terms of the processing and learning complexities. Furthermore, adopting the injection-based DPD allows to significantly reduce the LUT sizes, such that 16 or 32 entries are enough for efficient linearization, without interpolation. Additionally, the use of non-interpolated LUTs avoids the rank deficiencies in the SRA and sign-sign algorithms and thus the additional matrix transformation, which were experienced in [8]. Due to their low complexity and closed-loop nature, the developed solutions allow for fast real-time adaptation, and thus potentially on-chip implementations and continuous learning.

Extensive RF measurement results at 28 GHz, utilizing a state-of-the-art 64-element active antenna array and 400 MHz 5G NR waveform, are reported and analyzed. The obtained linearization results, together with the provided detailed complexity analysis, show that the proposed methods provide very favorable complexity-performance tradeoffs, while meeting the 3GPP 5G NR [4] emission requirements for FR2. Overall, the results show that efficient mmW array linearization can be obtained through the proposed methods.

$$\mathbf{\Phi} = \begin{bmatrix} x[n]\boldsymbol{\xi}_n^T & x[n-1]\boldsymbol{\xi}_{n-1}^T & \cdots & x[n-M]\boldsymbol{\xi}_{n-M}^T \\ x[n+1]\boldsymbol{\xi}_{n+1}^T & x[n]\boldsymbol{\xi}_n^T & \cdots & x[n-M+1]\boldsymbol{\xi}_{n-M+1}^T \\ \vdots & \vdots & \ddots & \vdots \\ x[n+N-1]\boldsymbol{\xi}_{n+N-1}^T & x[n+N-2]\boldsymbol{\xi}_{n+N-2}^T & \cdots & x[n+N-M-1]\boldsymbol{\xi}_{n+N-M-1}^T \end{bmatrix} \tag{1}$$

II. Closed-Loop DPD System

In this work, we adopt the MP DPD model, where the high-order polynomial functions are replaced with Q entry-sized LUTs [10]. This model is adopted due to its low processing complexity [9], [12]. Additionally, the system builds on a closed-loop learning architecture, where the DPD coefficients are directly adapted using the input signal $x[n]$ and observed signal $y[n]$, obtained from an over-the-air (OTA) observation receiver, following the notations shown in Fig. 1. It is also noted that even though our primary applications are in the mmW array transmitters, the proposed techniques are applicable to any single-input single-output DPD systems.

Formally, the input-output relation of the DPD, for a block of N samples, is expressed as

$$\mathbf{x}_{\text{DPD}} = \mathbf{x} + \mathbf{\Phi}\mathbf{w}, \tag{2}$$

where $\mathbf{\Phi} \in \mathbb{C}^{N\times C}$ is the input data matrix, whose structure in the MP case with memory depth M is shown in (1), $\mathbf{x} = [x[n], x[n+1], \cdots, x[n+N-1]]^T$ is the input data vector, and $C = MQ$ is the total number of model coefficients. The vector $\mathbf{w} \in \mathbb{C}^{C\times 1}$ contains the LUT entries, and it is initialized as a zero vector in the first DPD iteration. In (1), the vector $\boldsymbol{\xi}_n$ is defined as

$$\boldsymbol{\xi}_n \in \mathbb{R}^{Q\times 1} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T, \tag{3}$$

where the unit element is inserted in the index p_n, defined as

$$p_n = \left\lfloor \frac{|x[n]|}{\Delta_x} \right\rfloor + 1. \tag{4}$$

Thus, the input sample $x[n]$ is multiplied with the corresponding LUT gain, which is indexed by the input magnitude $|x[n]|$. Δ_x is the amplitude spacing of the LUT entries, defined as the maximum input magnitude divided by the desired number of LUT entries, Q.

Formulating the LUT-based DPD as a linear-in-parameters model as in (2), allows us to apply traditional closed-loop learning techniques, such as GN or LMS-type adaptations. Defining the error signal $\mathbf{e}_k \in \mathbb{C}^{N\times 1} = \mathbf{x}_k - \frac{\mathbf{y}_k}{G}$, for block iteration k, the damped GN and block-LMS learning rules can be defined, respectively, as [8], [10], [11]

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\big(\mathbf{\Phi}_k^H\mathbf{\Phi}_k\big)^{-1}\mathbf{\Phi}_k^H\mathbf{e}_k, \tag{5}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\mathbf{\Phi}_k^H\mathbf{e}_k, \tag{6}$$

where μ is the learning rate.

Finally, we note that the formulations in (2)–(6) are quite general, and can be applied with other LUT-based DPD models as well, such as those following generalized MP or Volterra-DDR models (see [10] for an example).

III. Signed Algorithms

The classical definition of the complex signum function projects a non-zero complex number to the unit circle of the complex plane [13]. The magnitude of the resulting number, $\bar{z}$, is 1, but the real and imaginary parts are not equal to ± 1, thus no complexity reduction can be achieved when multiplying with $\bar{z}$. To remove the need for multiplications, we define the complex signum function instead as

$$\text{csgn}(z) := \text{sgn}(\text{Re}(z)) + j\,\text{sgn}(\text{Im}(z)), \tag{7}$$

which provides either -1 or +1 for the real and imaginary parts. For matrices, the operation is taken element-wise.

A. The Sign Algorithm

The sign algorithm is obtained by signing the error signal $\mathbf{e}_n$ in the learning rules presented in (5) and (6). The motivation is to avoid multiplications in calculating $\mathbf{\Phi}^H\mathbf{e}$, such that only additions remain, which are less resource-intensive operations in digital signal processing (DSP) implementations. By signing the error vector, the DPD learning rules read

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\big(\mathbf{\Phi}_k^H\mathbf{\Phi}_k\big)^{-1}\mathbf{\Phi}_k^H\,\text{csgn}(\mathbf{e}_k), \tag{8}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\mathbf{\Phi}_k^H\,\text{csgn}(\mathbf{e}_k). \tag{9}$$

An implementation of the sign GN algorithm was shown in [7].

B. The Signed Regressor Algorithm

The SRA method signs the input data matrix in the learning rules. Multiplications in the terms $\mathbf{\Phi}^H\mathbf{\Phi}$ and $\mathbf{\Phi}^H\mathbf{e}$ (GN), and $\mathbf{\Phi}^H\mathbf{e}$ (LMS) are thus avoided, making the computational complexity of the learning rule lighter. The SRA learning rules can be expressed as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\Big(\text{csgn}(\mathbf{\Phi}_k^H)\mathbf{\Phi}_k\Big)^{-1}\,\text{csgn}(\mathbf{\Phi}_k^H)\mathbf{e}_k, \tag{10}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\,\text{csgn}(\mathbf{\Phi}_k^H)\mathbf{e}_k. \tag{11}$$

It is important to note that all polynomial-based DPD approaches, as well as linearly interpolated LUTs, will suffer from a rank deficiency in the signed data matrix $\text{csgn}(\mathbf{\Phi}_k^H)$, as repeated columns or linear combinations between them will appear. An example is presented in [8], in the context of an MP DPD. In such a case, the estimated DPD coefficients will diverge, as they do not have a unique solution. One way to solve this problem is to apply a unitary Walsh-Hadamard transformation to gaussianize the distribution of $\text{csgn}(\mathbf{\Phi}_k^H)$ and make it full rank [8]. This, however, further increases the complexity in the learning rule. On the other hand, with the proposed LUT-based DPD approach, the rank deficiencies are avoided, as the structure of this model does not lead to

Table 1. Complexity analysis of the normal and signed learning methods presented throughout the paper, in terms of real multiplications and real additions per DPD learning iteration. The last column presents a numerical example when $N = 25$ ksamples, $Q = 32$, $M = 4$, and $C = MQ = 128$.

	Real multiplications	**Real additions**	**Real mul. / Real Add.**
Gauss-Newton	$C^3 + 4M^2(N+1) + 2M(2N+1)$	$2\big(2M^2N + M(2N+M-1) + 2C\big)$	$(4\,/\,2) \times 10^6$
Sign Gauss-Newton	$C^3 + 4M^2(N+1) + 2C$	$4M^2N + 2M(N+M-2) + 2C + 2N\log_2(N)$	$(3.7\,/\,2.5) \times 10^6$
SRA Gauss-Newton	$C^3 + 4M^2 + 2M$	$2\big(M^2N + M(M+N-2) + 2C\big) + 2N\log_2\big(N\frac{M}{2}\big)$	$(2\,/\,1.8) \times 10^6$
Sign-sign Gauss-Newton	$C^3 + 2M$	$2\big(M^2(N-1) + M(N+M-2) + C\big) + 2N\log_2\big(N^2\frac{M}{2}\big)$	$(2\,/2.5\,) \times 10^6$
Block-LMS	$2M(2N+1)$	$2(MN+C)$	$(400\,/\,200) \times 10^3$
Sign block-LMS	$2M$	$M(N-1) + 2C + 2N\log_2(N)$	$8\,/\,100 \times 10^3$
SRA block-LMS	$2M$	$M(N-1) + 2C + 2N\log_2\big(N\frac{M}{2}\big)$	$8\,/\,880 \times 10^3$
Sign-sign block-LMS	0	$2M(N-1) + 2C + 2N\log_2\big(N^2\frac{M}{2}\big)$	$0\,/\,170 \times 10^3$

repeated or linearly dependent columns in $\mathrm{csgn}(\mathbf{\Phi}_k^H)$. Thus, the SRA learning rule can be directly applied, with no extra matrix transformations needed.

C. The Sign-Sign Algorithm

Finally, the sign-sign algorithm applies the signum function to both the data matrix and the error vector, further reducing the overall complexity. The same discussion about the rank deficiency problem applies here as well. The learning expressions with the sign-sign algorithm read

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\Big(\mathrm{csgn}(\mathbf{\Phi}_k^H)\mathbf{\Phi}_k\Big)^{-1}\mathrm{csgn}(\mathbf{\Phi}_k^H)\,\mathrm{csgn}(\mathbf{e}_k), \tag{12}$$

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu\,\mathrm{csgn}(\mathbf{\Phi}_k^H)\,\mathrm{csgn}(\mathbf{e}_k). \tag{13}$$

D. Learning Complexity Comparison

The learning complexity is analyzed in terms of real multiplications and real additions per DPD coefficient update, over an N-sized block of samples. It is assumed that one complex multiplication is implemented with 4 real multiplications and 2 real additions. Table 1 presents the complexity expressions of the GN and block-LMS adaptive learning methods, covering the original learning rules in (5) and (6) and the sign-based versions in (8)–(13). The last column shows a numerical example with $N = 25$ ksamples, $Q = 32$, $M = 4$, and $C = MQ = 128$, which is the same parameterization used in the experimental results in Section IV. The complexities of the signed GN algorithms are clearly reduced, but are still clearly higher than those of the LMS algorithms, mainly due to the required matrix inversion. The signed LMS algorithms provide remarkably simple options for learning.

IV. Experimental Results

A. 28 GHz Active Array Experimental Setup

Shifting towards mmW frequencies [5], all the experiments are carried out using an FR2 measurement setup, depicted in Fig. 2. The transmit chain consists of a Keysight M8190 waveform generator providing the I/Q samples at 3.5 GHz IF, further devices to upconvert the signal to 28 GHz, and finally an Anokiwave AWMF-0129 active antenna array. The signal is measured OTA, downconverted again to IF, and digitized for

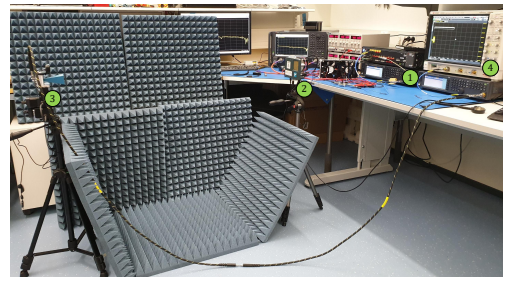


Fig. 2. RF measurement setup including the Keysight M8190 waveform generator (1), Anokiwave AWMF-0129 active antenna array (2), horn antenna as receiver (3), and the Keysight DSOS804A digitizer (4).

DPD processing. The adopted signals are NR FR2 compliant OFDM waveforms, with bandwidths of 100 and 400 MHz and subcarrier spacing of 60 and 120 kHz, respectively. The MP-LUT DPD models utilize LUT entry sizes of $Q = 32$ and $M = 4$ memory branches. The classical MP model ($P = 9$, $M = 4$) in a closed-loop configuration is used as a reference. The DPD operation is block-based, with a block size of 25 ksamples and 15 closed-loop iterations.

B. Measurement Results

In this section, two evaluations of the proposed DPD models are provided. The first test features two sets of OTA measurements using the 400 MHz signal, measured at a highly nonlinear operation point with EIRP ≈ 43 dBm. The measured PSDs are depicted in Fig. 3 a) and Fig. 3 b). In both cases, the performances of the sign and SRA algorithms are not degraded drastically in comparison with the normal learning methods, which lie at the same time close to the reference MP model. The performance of the sign-sign algorithm is also similar when GN-based learning is utilized, however, it is slightly degraded on the right hand side of the spectra when adopting LMS-type adaptation.

The second test, presented in Fig. 4, features an OTA measurement with 100 MHz NR signal and a varying LUT entry size Q, while EIRP ≈ 41 dBm. It is seen from the figure that increasing Q up to 32 improves the DPD linearization performance, which basically saturates when Q is further

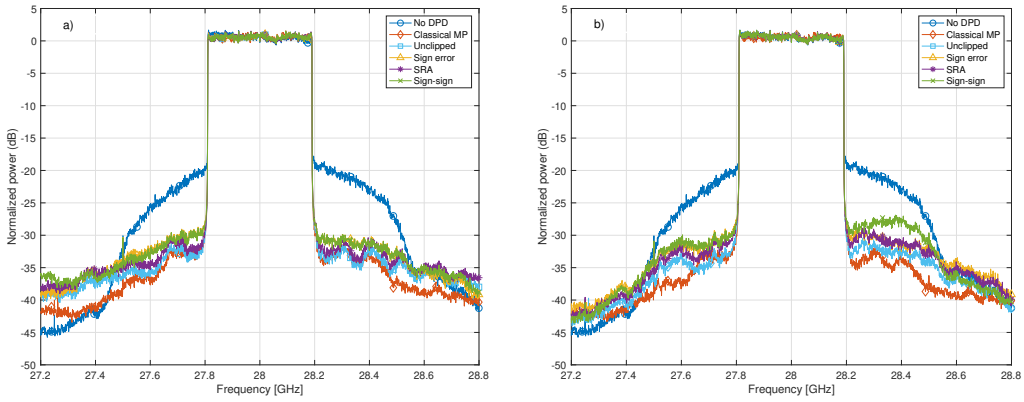


Fig. 3. OTA NR FR2 400 MHz linearization performance at EIRP $\approx$ 43 dBm, with original and signed algorithms with a) damped GN and b) block-LMS.

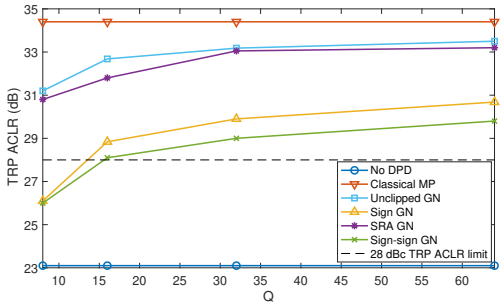


Fig. 4. OTA NR FR2 100 MHz linearization performance at EIRP $\approx$ 41 dBm, when varying the LUT entry size, $Q = 8, 16, 32, 64$.

increased to 64. Compared to the results in [10], the injection-based DPD structure seems to allow for lower entry sized non-interpolated LUTs, while the sign algorithms further reduce the processing and learning complexities. The linearization performances lie only 0.2 dB (SRA) and 3-3.5 dB (sign, sign-sign) from the unclipped MP-LUT model. The 5G NR ACLR limit of 28 dBc, measured using the total radiated power (TRP) [4] approach, is fulfilled in all cases except when considering the sign and sign-sign algorithms with $Q = 8$.

V. Conclusions

In this paper, we formulated various signed closed-loop DPD learning algorithms for LUT-based memory DPD, assuming the injection-based DPD structure. Due to the injection-based DPD design, the LUT entry size required in the DPD models was decreased, such that 16 or 32 entries were enough for efficient linearization. Additionally, the use of LUTs avoided rank deficiencies in the SRA and sign-sign algorithms, thus eliminating the need for additional matrix transformations required by earlier approaches. Extensive measurements using a state-of-the-art 28 GHz active antenna array and up to 400 MHz 5G NR waveforms were reported to validate the techniques. These results, together with a complexity analysis, demonstrate that the proposed models have a very favorable complexity-performance tradeoff.

Acknowledgment

This work was financially supported by the Academy of Finland under the projects 301820, 323461 and 319994.

References

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

[2] R. N. Braithwaite, "General principles and design overview of digital predistortion," in *Digital Processing for Front End in Wireless Communication and Broadcasting*. Cambridge Univ. Press, 2011, ch. 6, pp. 143–191.

[3] A. S. Tehrani *et al.*, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Tech.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.

[4] 3GPP Tech. Spec. 38.104, "NR; Base Station (BS) radio transmission and reception," v15.4.0 (Release 15), Dec. 2018.

[5] C. Fager *et al.*, "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.

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

[7] H. Wang *et al.*, "1-bit observation for direct-learning-based digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2465–2475, Jul. 2017.

[8] N. Guan, N. Wu, and H. Wang, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 921–923, Oct 2018.

[9] Y. Ma, Y. Yamao, Y. Akaiwa, and C. Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 11, pp. 3961–3973, Nov. 2013.

[10] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980–987, Nov. 2017.

[11] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 280–293, Jan. 2018.

[12] C. D. Presti *et al.*, "Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA PA module," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 3, pp. 604–618, 2012.

[13] J. A. Tropp, "Recovery of short, complex linear combinations via l1 minimization," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1568–1570, Apr. 2005.

PUBLICATION 7

Adaptive cancellation of nonlinear self-interference in wireless full-duplex: cascaded spline-interpolated methods

P. Pascual Campo, L. Anttila, D. Korpi and M. Valkama

2020 54th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA 2020, 1265–1271

DOI: 10.1109/IEEECONF51394.2020.9443406

Publication reprinted with the permission of the copyright holders

Adaptive Cancellation of Nonlinear Self-Interference in Wireless Full-Duplex: Cascaded Spline-Interpolated Methods

Pablo Pascual Campo, *Graduate Student Member, IEEE*, Lauri Anttila, *Member, IEEE*, Dani Korpi, *Member, IEEE*, and Mikko Valkama, *Senior Member, IEEE*

***Abstract*—This article studies advanced digital self-interference (SI) cancellation algorithms for wireless full-duplex (FD) transceivers. While the majority of works in this area consider linear SI models and corresponding cancellation algorithms, we take also both the transmitter (TX) and receiver (RX) nonlinearities into account. Specifically, we develop a Hammerstein-Wiener type of digital SI cancellation system where the two involved nonlinearities are efficiently implemented through spline-interpolated look-up tables (LUTs). Also, adaptive parameter estimation solutions are pursued and discussed, and comprehensive cancellation performance examples and processing complexity comparisons are provided. The obtained results show that taking both the TX and RX nonlinearities into account in the digital cancellation stage can be beneficial, particularly when the TX power amplifier is operating close to saturation and when the RF isolation and RF cancellation prior to the RX low-noise amplifier (LNA) is limited, such that also the LNA is distorting the SI waveform.**

***Index Terms*—Self-interference, spline interpolation, in-band full-duplex, cancellation, look-up table, adaptive, nonlinear distortion**

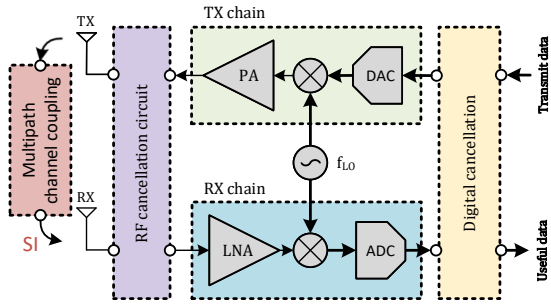


Fig. 1: Simplified generic full-duplex transceiver block diagram, showing the main sources contributing to the SI.

I. Introduction

The potential of full-duplex technology has been demonstrated by several research groups [1]–[5], but much more implementation-oriented research is needed to fulfill the potential. In theory, full-duplex is capable of doubling the spectral efficiency of a wireless communication system by transmitting and receiving information simultaneously at the same frequency channel. The resulting data rate increase is particularly sought within the heavily congested ultra high frequency (UHF) radio frequency spectrum, since full-duplex operation does not require any additional bandwidth.

Although full-duplex may be a key technology for future wireless networks, also new challenges appear. A particular challenge is the inevitable self-interference (SI), which can be more than 100 dB stronger than the weak desired signal at the receiver input [6]. The SI is caused by the own transmit signal which is leaked back to the full-duplex receiver, overlapping the desired received signal both temporally and spectrally. Any full-duplex device must be capable of suppressing it in order to operate properly, and many coexisting techniques are typically used, as shown in the transceiver block diagram in Fig. 1. The digital cancellation stage is the last cancellation stage, with the aim of reducing the SI level below the noise floor, by accurately modeling and accounting for the coupling channel effects [5], [7]. Since any full-duplex device must perform the SI cancellation, it is key to explore ways of minimizing the involved computational complexity, while at the same time provide the required SI cancellation.

In this paper, we propose a Hammerstein-Wiener type of digital SI cancellation system which utilizes spline-interpolated look-up tables (LUTs) [7], [8] to efficiently model the two involved nonlinearities in the cascaded model. Both nonlinearities aim at modelling the transmit (TX) power amplifier (PA) and the receiver (RX) low noise amplifier (LNA), thus taking into account both the TX and RX path effects. Spline modeling allows for significant complexity reduction when compared to classical models, as already discussed in our earlier work [7]. Additionally, adaptive parameter estimation solutions, based on the steepest gradient-descent, are derived and discussed for each involved parameter. A detailed complexity analysis is also presented to assess the complexity-performance trade-offs of the presented models.

Then, extensive RF cancellation results at 28 GHz (5G NR band n257 [9]), utilizing a state-of-the-art 64-element active antenna array and 5G NR like OFDM waveforms, are reported and analyzed, incorporating standard-compliant channel bandwidths of 100 and 400 MHz. The classical Wiener and Hammerstein models, where the instantaneous nonlinearity is also implemented with spline-interpolated LUTs, are also measured and reported as a reference. The obtained cancellation results, together with the provided detailed complexity analysis, demonstrate the favorable performance-complexity

trade-off of the proposed model.

The rest of this article is organized as follows. In Section II, the proposed Hammerstein-Wiener model is described, along with the derived adaptive parameter update learning equations. Section III presents a detailed complexity analysis of the proposed algorithm, in terms of real multiplications per sample. After this, the cancellation performance of the proposed and reference models is evaluated and compared in Section IV, with real RF measurements. Finally, Section V concludes and summarizes the main findings of this paper.

Notation used in this paper

In this paper, vectors are represented with boldface lowercase letters. By default, all vectors consist of complex-valued elements represented as columns vectors (i.e., $\mathbf{x} \in \mathbb{C}^{N\times 1} = [x_0 \; x_1 \; \cdots \; x_{N-1}]^T$). Matrices are expressed with boldface capital letters (i.e., $\mathbf{A} \in \mathbb{C}^{N\times M}$). Ordinary transpose of vectors is represented as $(\cdot)^T$. Moreover, the complex conjugate, absolute value, and floor operator are denoted by $(\cdot)$, $|\cdot|$, and $\lfloor\cdot\rfloor$, respectively.

II. Spline-Based Hammerstein-Wiener Model

In this section, we study the spline-based Hammerstein-Wiener model (denoted here as SPHW), a block-oriented system that serially connects elementary linear and nonlinear blocks to ideally model a system whose structure is alike [10], [11]. The SPHW method has two nonlinear blocks and a linear finite impulse response (FIR) filter block connected in between, as illustrated in Fig. 2. In this work, the nonlinear blocks are implemented with an injection-based uniform spline interpolated LUT, and the linear block is implemented as an FIR filter. In the context of in-band full-duplex, and following the nomenclature used in Fig. 1, the LUT-based nonlinear elementary blocks model the nonlinear responses of the TX PA and the RX LNA, and the FIR filter models the multipath channel response. The output signal of the model is then the estimated SI signal, denoted as $y[n]$. The digital SI canceller (DSIC) will subtract this perturbation from the received signal, $d[n]$, in order to suppress its interfering effect. The SPHW structure and the spline interpolated LUTs are adopted due to its inherent low complexity, while they still achieve excellent levels of SI cancellation, as demonstrated in later sections.

A. SI Regeneration and Cancellation

Following the serial order described above, and making reference to Fig. 2, the regenerated SI signal, $y[n]$, is obtained as follows.

First, a preceding nonlinear block is considered, providing the output signal $l[n]$. The spline interpolated LUT scheme divides the input magnitude, $|x[n]|$, in C regions. Each region is modelled with a spline polynomial of order P. The characterization of the regions is given through the index, abscissa, and abscissa vector local parameters, which constitute the basis for the spline interpolation scheme [12], and read

$$i_n = \left\lfloor \frac{|x[n]|}{\Delta_{\mathrm{x}}} \right\rfloor + 1, \tag{1}$$

$$u_n = \frac{|x[n]|}{\Delta_{\mathrm{x}}} - (i_n - 1), \tag{2}$$

$$\mathbf{u}_n \in \mathbb{R}^{(P+1)\times 1} = \begin{bmatrix} u_n^P & u_n^{P-1} & \cdots & 1 \end{bmatrix}^T, \tag{3}$$

where Δ_{x} represents the region spacing. According to this definition, i_n corresponds to the region number, and u_n is the normalized value within each spline region.

Once the region parameters are characterized from the input sample, the injection-based interpolation scheme can be applied to obtain the output value, $l[n]$, as

$$l[n] = x[n]\mathbf{\Psi}_n^T(\mathbf{1} + \mathbf{c}_n). \tag{4}$$

Here, $\mathbf{1} \in \mathbb{R}^{C\times 1}$ is a vector of all ones, $\mathbf{c}_n \in \mathbb{C}^{C\times 1}$ contains the complex LUT control points used in the interpolation scheme, and the vector $\mathbf{\Psi}_n \in \mathbb{R}^{C\times 1}$ contains nonzero values starting from index i_n, such that only the appropriate control points are selected as weights for the interpolation scheme (i.e. $(c_{i_n} \, c_{i_n+1} \cdots c_{i_n+P})$). It is formally defined as

$$\mathbf{\Psi}_n \in \mathbb{R}^{C\times 1} = [0 \; \cdots \; 0 \; \mathbf{u}_n^T\mathbf{B}_P \; 0 \; \cdots \; 0]^T, \tag{5}$$

where $\mathbf{B}_P \in \mathbb{R}^{(P+1)\times(P+1)}$ corresponds to the *spline basis matrix*, which depends on the considered spline order, and can be found in, for instance, [8], [13].

It is worth noting that the injection-based scheme, in combination with the spline interpolated LUT, allows for two main things. First, the gain ambiguities between the linear and nonlinear cascaded blocks are effectively removed. Second, the dynamic range of the LUT control point vector is reduced, requiring less bits and thus making fixed-point implementations more efficient [14].

The signal $l[n]$ is then carried through the linear block, implemented as a FIR filter. Its output can be directly written as

$$s[n] = \mathbf{w}_n^T\mathbf{l}_n, \tag{6}$$

where $\mathbf{w}_n \in \mathbb{C}^{M\times 1} = [w_0 \; w_1 \; \cdots \; w_{M-1}]^T$ are the FIR filter coefficients, and $\mathbf{l}_n \in \mathbb{C}^{M\times 1} = [s[n] \; s[n-1] \; \cdots \; s[n-M+1]]^T$ is the signal regression of $l[n]$.

Finally, the obtained signal $s[n]$ is passed through the second nonlinear block. The process is the same as presented above, but considering now a spline order K, and number of created regions Q. Mimicking the expressions presented in (1), (2), (3), (5), but considering $s[n]$ as the input signal, the interpolation scheme is built, and the model output signal, $y[n]$, is obtained, as

$$y[n] = s[n]\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n), \tag{7}$$

where $\mathbf{q}_n \in \mathbb{C}^{Q\times 1}$ corresponds to the second LUT control point vector, and $\mathbf{\Phi}_n \in \mathbb{R}^{Q\times 1}$ is defined in the same way as in (5).

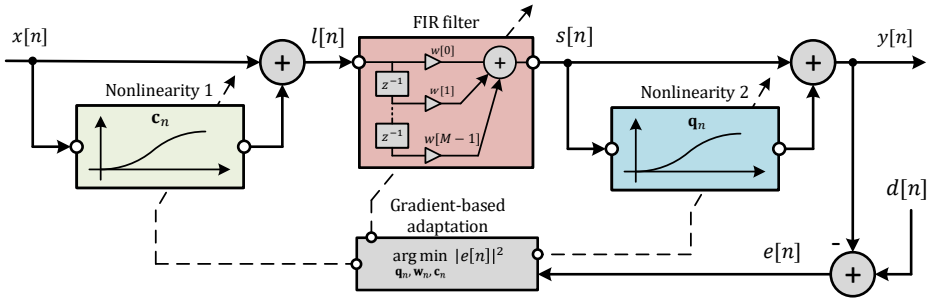


Fig. 2: Block diagram of the proposed spline-based Hammerstein-Wiener system for identifying and cancelling the SI signal.

It is noted that different parametrization (spline order and number of regions) can be used in each individual nonlinearity. This can be useful if one of the amplifiers exhibits deeper nonlinear effects than the other. The parametrization can be consequently increased or decreased accordingly to optimize performance and/or complexity.

Finally, once the SI signal $y[n]$ is regenerated, it can be suppressed in the final digital cancellation stage by subtracting it from the received signal (i.e., $d[n] - y[n]$).

B. Learning Updates

In order to estimate and track the involved parameters in the SPHW unknown system, three different learning rules are derived to adapt each of the considered parameters (i.e. $\mathbf{c}_n$, $\mathbf{w}_n$, and $\mathbf{q}_n$). Firstly, let us denote the error signal used in the estimation as

$$e[n] = d[n] - y[n], \tag{8}$$

where $d[n]$ denotes the received signal the digital canceller aims at suppressing, and $y[n]$ is the SI regeneration obtained in the previous section. At this point, the problem lies in obtaining the mathematical formulation that minimizes $e[n]$, for each parameter. To this end, we utilize the least mean square (LMS) gradient-based adaptation, in which the parameters are adapted following the negative steepest descent direction of the cost function, given by the gradient. The cost function depending on $\mathbf{c}_n$, $\mathbf{w}_n$, $\mathbf{q}_n$, can be defined, in turn, as

$$J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n) = e[n]e^*[n]. \tag{9}$$

At this point, three assumptions can be made regarding the cascaded structure of the SPHW model:

- The first nonlinearity, $\mathbf{c}_n$, depends on the FIR filter and second nonlinearity.
- The FIR filter, $\mathbf{w}_n$, depends on the second nonlinearity, but not on the first.
- The second nonlinearity, $\mathbf{q}_n$, does not depend on the first nonlinearity nor the FIR filter.

Taking these assumptions into account, the general form of the learning rule for the first nonlinearity, $\mathbf{c}_n$, can be written as

$$\mathbf{c}_{n+1} = \mathbf{c}_n - \mu_{\mathrm{c}} \nabla_{\mathbf{c}_n} J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n), \tag{10}$$

where μ_{c} is the learning rate for the update, and $\nabla_{\cdot}(\cdot)$ refers to the complex gradient operator. Making use of elementary differentiation rules, the derivative with respect $\mathbf{c}_n^*$, keeping the other parameters fixed, can be calculated as

$$\begin{aligned}
\frac{\partial J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{c}_n^*} &= e[n]\frac{\partial e^*[n]}{\partial \mathbf{c}_n^*} + e^*[n]\frac{\partial e[n]}{\partial \mathbf{c}_n^*} \\
&= -e[n]\frac{\partial s^*[n]\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*)}{\partial \mathbf{c}_n^*} + 0 \\
&= -e[n]\frac{\partial \mathbf{l}_n^H}{\partial \mathbf{c}_n^*}\mathbf{w}_n^*\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*) \\
&= -e[n]\mathbf{\Sigma}_n^T\mathbf{X}_n^*\mathbf{w}_n^*\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*),
\end{aligned} \tag{11}$$

where $\mathbf{X}_n \in \mathbb{C}^{M\times M} = \mathrm{diag}(x[n], x[n-1], \cdots, x[n-M+1])$, and $\mathbf{\Sigma}_n = (\mathbf{\Psi}_n\mathbf{\Psi}_{n-1}\cdots\mathbf{\Psi}_{n-M+1})^T$. Additionally, it is assumed that the rate of change of $\mathbf{c}_n$ over the span of the filter length M is negligible, i.e., $\mathbf{c}_n \approx \mathbf{c}_{n+M}$. This assumption is made considering the small value of μ_{c}.

Therefore, the final expression for the learning rule of $\mathbf{c}_n$ reads

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \mu_{\mathrm{c}} e[n]\mathbf{\Sigma}_n^T\mathbf{X}_n^*\mathbf{w}_n^*\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*). \tag{12}$$

Secondly, the learning rule for the FIR filter, $\mathbf{w}_n$, can be expressed as

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu_{\mathrm{w}} \nabla_{\mathbf{w}_n} J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n), \tag{13}$$

where μ_{w} is the learning rate. Again, by obtaining the deriva-

TABLE I: Computational complexity in terms of real multiplications per sample of the SPHW model, as a function of the modelling parameters P, K M, and τ. In the third column, a numerical example is presented, when $P = 3$, $K = 3$, $M = 76$, and $\tau = 5$.

Operation		Symbolic real multiplications	Numerical real multiplications
Model identification	$l[n]$	$P^2 + 4P + 8 + sqrt$	29
	$s[n]$	$4M$	304
	$y[n]$	$K^2 + 4K + 8 + sqrt$	29
	Total	$P^2 + K^2 + 4P + 4K + 4M + 16 + 2sqrt$	362
Coefficient update	$\mathbf{q}_{n+1}$	$2K + 8$	14
	$\mathbf{w}_{n+1}$	$K^2 + 3K + 6M + 15 + div$	489
	$\mathbf{c}_{n+1}$	$4P + 2P\tau + 6\tau + 10$	82
	Total	$4P + 2P\tau + 6\tau + K^2 + 5K + 6M + 33 + div$	585
Total per iteration		$P^2 + 8P + 2P\tau + 6\tau + 2K^2 + 9K + 10M + 49 + div + 2sqrt$	947

tive with respect to $\mathbf{w}_n^*$, we obtain

$$\begin{aligned}\frac{\partial J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{w}_n^*} &= e[n]\frac{\partial e^*[n]}{\partial \mathbf{w}_n^*} + e^*[n]\frac{\partial e[n]}{\partial \mathbf{w}_n^*} \\ &= -e[n]\frac{\partial \mathbf{w}_n^H \mathbf{l}_n^* \mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*)}{\partial \mathbf{w}_n^*} \\ &\quad - e^*[n]\frac{\partial \mathbf{w}_n^T \mathbf{l}_n \mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n)}{\partial \mathbf{w}_n^*} \\ &= -e[n]\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*)\mathbf{l}_n^* - \frac{\mathbf{l}_n^* s[n] \operatorname{Re}\{e[n]\dot{y}[n]\}}{\Delta_{\mathrm{s}}|s[n]|},\end{aligned} \tag{14}$$

where $\dot{y}[n] = s^*[n]\dot{\mathbf{u}}_n^T \mathbf{B}_K(\mathbf{1} + \mathbf{q}_{n,i_n}^*)$ and the vector $\mathbf{q}_{n,i_n}$ makes reference to the column $\mathbf{q}_{n,i_n} \in \mathbb{C}^{(K+1)\times 1} = [q_{i_n} \cdots q_{i_n+K}]^T$.

Thus, the final expression for the learning update of the FIR filter, $\mathbf{w}_n$, reads

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu_{\mathrm{w}} \mathbf{l}_n^* \Big(e[n]\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*) + \frac{s[n] \operatorname{Re}\{e[n]\dot{y}[n]\}}{\Delta_s |s[n]|} \Big). \tag{15}$$

Finally, the learning rule for the second nonlinearity, $\mathbf{c}_n$, can be expressed as

$$\mathbf{q}_{n+1} = \mathbf{q}_n - \mu_{\mathrm{q}} \nabla_{\mathbf{q}_n} J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n), \tag{16}$$

where μ_{q} is the learning rate. Differentiating with respect $\mathbf{q}_n^*$, we obtain

$$\begin{aligned}\frac{\partial J(\mathbf{c}_n, \mathbf{w}_n, \mathbf{q}_n)}{\partial \mathbf{q}_n^*} &= e[n]\frac{\partial e^*[n]}{\partial \mathbf{q}_n^*} + e^*[n]\frac{\partial e[n]}{\partial \mathbf{q}_n^*} \\ &= -e[n]\frac{\partial s^*[n]\mathbf{\Phi}_n^T(\mathbf{1} + \mathbf{q}_n^*)}{\partial \mathbf{c}_n^*} + 0 \\ &= -e[n]s^*[n]\mathbf{\Phi}_n.\end{aligned} \tag{17}$$

And thus, the final learning rule reads

$$\mathbf{q}_{n+1} = \mathbf{q}_n + \mu_{\mathrm{q}} e[n] s^*[n] \mathbf{\Phi}_n. \tag{18}$$

These learning equations can thus be executed iteratively to track possible dynamic changes in the system, or kept fixed if the system is expected to remain static.

III. Complexity Analysis

In this section, we study the computational complexity involved in the proposed SPHW method. Together with the experimental results presented in the next section, it will allow to assess the performance-complexity trade-off of the SPHW model.

In this paper, the computational complexity is divided in two stages, cancellation and coefficient update, and it is presented in terms of real multiplications per sample. Real multiplications constitute a critical metric for digital signal processing (DSP) implementations, while additions are essentially free [15]. For the analysis, it is assumed that one complex multiplications costs 4 real multiplications, and one complex-real multiplication is calculated with 2 real multiplications.

Table I presents the symbolic computational complexity expressions as a function of the spline order, P (first nonlinear block), and K (second nonlinear block), and as a function of the FIR filter memory taps, M. Additionally, a complexity reduction method in the learning update of $\mathbf{c}_n$ is proposed in order to ease the coefficient update. This method is formulated by considering the additional parameter τ, which truncates the time dimension of $\mathbf{\Sigma}_n$ (row dimension), by only considering a specific number of past samples in the update. This process is motivated as follows. $\mathbf{\Sigma}_n$ is usually a big matrix if M is large, however, the most significant memory taps contributing to the learning update are usually the most recent ones, thus by selecting only these, the computational complexity can be greatly reduced, while the cancellation performance is almost not affected. Note that if $\tau = M$, no approximation is done.

Finally, the second column of Table I shows the exact numerical complexity as an specific example, when the spline order is considered as $P = 3$, $K = 3$, the filter memory is considered as $M = 76$, and the complexity reduction coefficient is chosen as $\tau = 5$. This parameterization is the same as the one used for validating the results in the next section.

IV. Experimental Results

This section presents the RF measurement setup utilized to evaluate the performance of the SPHW method, along with the corresponding obtained RF SI cancellation results.

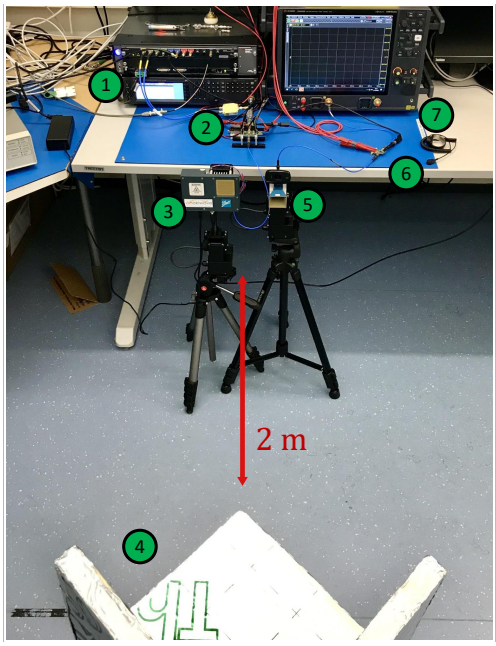


Fig. 3: RF measurement setup including the Keysight M8195A AWG and Keysight N5183B-MXG LO signal generator (1), upconversion to 28 GHz (NR Band n257) and pre-amplifying stage (2), Anokiwave AWMF-0129 active antenna array (3), metallic reflector to generate the coupling SI (4), horn antenna as receiver (5), AD HMC1040LP3CE LNA (6), and Keysight UXR0402AP digitizer (7).

A. Measurement Setup

In order to evaluate the cancellation performance of the proposed SPHW, the mmW RF test setup presented in Fig. 3 has been utilized. The proposed setup consist on the following elements. First, and arbitrary waveform generator (AWG) is utilized to generate the transmit I/Q data, which is then upconverted to 28 GHz and pre-amplified in order to facilitate a sufficiently high saturation power in the TX array. Then, the signal is transmitted over-the-air (OTA) by an Anokiwave AWMF-0219, which is a 64-element active antenna array transmitter. A metallic reflector was placed at 2 meters from the system to generate a possible environmental reflection, which is leaking to the FD receiver causing the SI. Finally, the received OTA signal is captured by a horn antenna, driven through a LNA, and inputted back to the digitizer in order to facilitate the corresponding post-processing in a host PC, where the DSIC algorithms are executed.

B. BW 100 MHz Digital Cancellation Results

This section evaluates the digital cancellation performance over a signal bandwidth of 100 MHz. The generated signal is a 3GPP 5G new radio (NR) Release-15 frequency range 2 (FR-2) compliant OFDM waveform, with 120 kHz subcarrier spacing (SCS) and 66 allocated resource blocks (RBs). This configuration maps to the aforementioned signal bandwidth of 100 MHz [9]. Additionally, the initial peak-to-average power ratio (PAPR) of the digital waveform is 9.5 dB, when measured at the 0.01 % point of the instantaneous PAPR complementary cumulative distribution function (CCDF), and is then limited to 7 dB through iterative clipping and filtering. An additional time-domain window is also applied to suppress the inherent OFDM signal sidelobes.

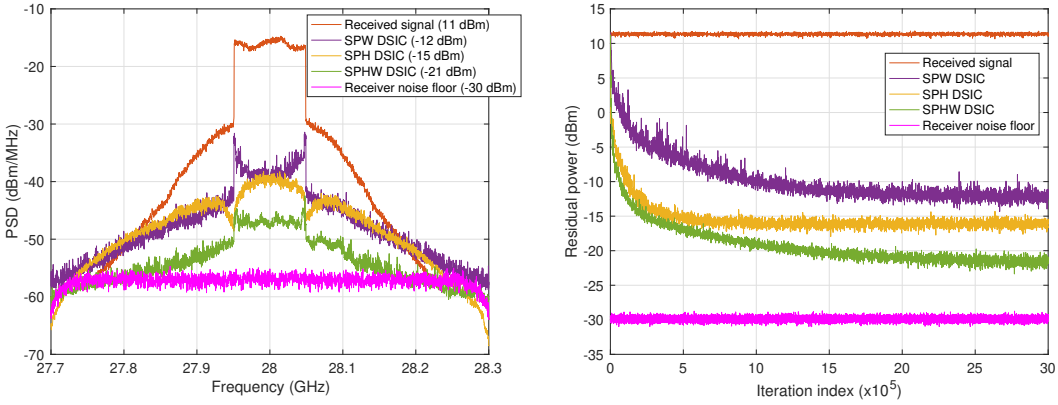
The parametrization chosen for the SPHW model is $P = K = 3$, $C = Q = 7$ for the spline order and LUT size, and $M_{\text{pre}}/M_{\text{post}} = 25/50$ for the pre-cursor and post-cursor memory taps in the FIR filter. Additionally, the spline-based Wiener (SPW) and Hammerstein (SPH) models, configured with $K = 3$, $Q = 7$ (SPW), $P = 3$, $C = 7$ (SPH), and $M_{\text{pre}}/M_{\text{post}} = 25/50$, have also been measured and presented for reference. The experimental results are then presented in Fig. 4, and summarized in Table II. It can be seen that the best SI cancellation performance is obtained with the proposed SPHW model (32 dB), followed by the SPH model (26 dB), and finally by the SPW model (22 dB). This result can be explained as follows. The SPHW approach has richer modeling capabilities than the other two models, as it includes two instantaneous nonlinear functions in its cascaded structure, which successfully model the behavior of both TX PA and RX LNA. Secondly, the inherent cascaded structure of the SPH approach is able to successfully model the TX antenna array, which is the predominant source of nonlinear distortion in this system. Finally, the SPW approach is not fully capable of modeling the distortion injected by the antenna array, due to its cascaded structure, thus providing a somewhat more degraded cancellation performance.

Finally, it can be seen from Fig. 4b that the convergence speed is reduced when considering the SPHW approach, compared to SPW or SPH. The former model has an increased number of parameters to be estimated, thus it takes more iterations to reach the final steady-state.

C. BW 400 MHz Digital Cancellation Results

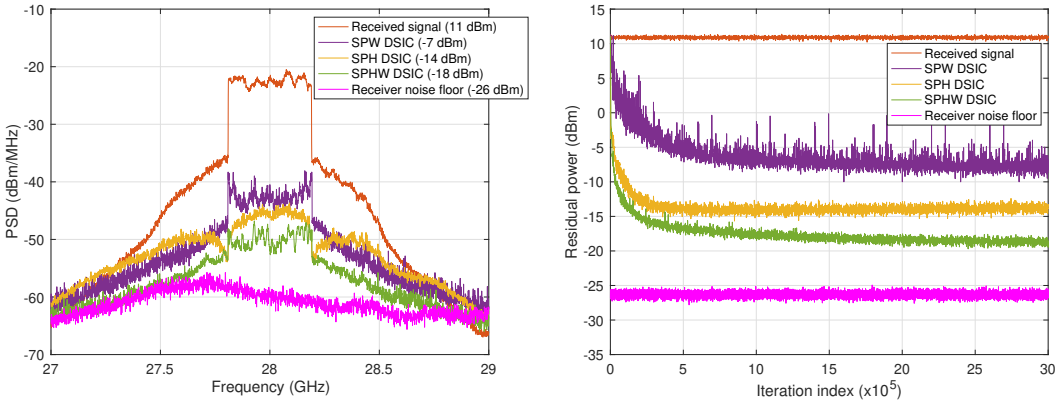
The second experiment further pushes the performance boundaries by considering an increased signal bandwidth of 400 MHz. The generated FR-2 NR signal has now 264 allocated RBs, while the SCS and PAPR remain the same as before, thus yielding a bandwidth of 400 MHz [9].

In this experiment, the parametrization chosen for the SPHW model is $P = K = 3$, $C = Q = 7$ for the spline order and LUT size, and $M_{\text{pre}}/M_{\text{post}} = 25/60$ for the pre-cursor and post-cursor memory taps in the FIR filter. The SPW and SPH models have also been measured for reference, configured with $K = 3$, $Q = 7$ (SPW), $P = 3$, $C = 7$ (SPH), and $M_{\text{pre}}/M_{\text{post}} = 25/60$. The overall memory of the models have been increased to better model the wider SI coupling channel. The obtained results are then presented in Fig. 5, and summarized in Table III. The proposed SPHW model is capable of obtaining the best SI cancellation (29 dB), followed by the SPH approach (25 dB), and finally by



(a) (b)

Fig. 4: Power spectral densities in (a) and residual power with respect to the iteration index in (b) of the overall signal after the different digital cancellers for 100 MHz signal bandwidth.



(a) (b)

Fig. 5: Power spectral densities in (a) and residual power with respect to the iteration index in (b) of the overall signal after the different digital cancellers for 400 MHz signal bandwidth.

the SPW scheme (18 dB). Similar observations to those of the previous experiment can be drawn. Since the SPHW model takes into account both TX and RX nonlinearities, it is capable of achieving an enhanced cancellation performance, while the SPW and SPH models, only considering either the TX or RX nonlinearity, lie somewhat behind.

It is also noted that the convergence speed of the SPHW approach is somewhat slower compared to SPW or SPH, as seen from Fig. 5b, since the model has a greater number of parameters to be updated.

V. Conclusions

In this article, we proposed a novel digital self-interference canceller for in-band full-duplex devices. The proposed model was a Hammerstein-Wiener type of system such that both TX and RX nonlinearities can be taken into account. Specifically, the instantaneous nonlinearities in the model were implemented through spline-interpolated look-up tables. Efficient adaptive parameter estimation solutions, based on the steepest gradient-descent algorithm, were also presented and discussed. Moreover, a detailed complexity analysis of the proposed model was reported, in terms of real multiplications per sample. Then, comprehensive RF measurements showing the cancellation performance of the proposed model were presented and compared to other existing solutions. The obtained results showed that taking both TX and RX nonlinearities into account can be beneficial, particularly when the TX power amplifier is operated close to saturation, and when the RF isolation and RF cancellation cannot provide sufficient SI cancellation, such

TABLE II: The RF measurement and model parameters in the DSIC experiment 1. The complexity in the cancellation and parameter update stages is presented in the last two columns, in terms of real multiplications per sample.

System parameters	**Value**	**Model**	P/C	$M_{\mathrm{pre}}/M_{\mathrm{post}}$	K/Q	**Canc.**	**Mul. - canc.**	**Mul. - update**
Transmit waveform	100 MHz NR OFDM @ 28 GHz	SPW DSIC	- / -	25 / 50	3 / 7	23 dB	333	503
EIRP	43 dBm	SPH DSIC	3 / 7	25 / 50	- / -	26 dB	333	384
Received power ($d[n]$)	11 dBm	SPHW DSIC	3 / 7	25 / 50	3 / 7	32 dB	362	585

TABLE III: The RF measurement and model parameters in the DSIC experiment 2. The complexity in the cancellation and parameter update stages is presented in the last two columns, in terms of real multiplications per sample.

System parameters	**Value**	**Model**	P/C	$M_{\mathrm{pre}}/M_{\mathrm{post}}$	K/Q	**Canc.**	**Mul. - canc.**	**Mul. - update**
Transmit waveform	400 MHz NR OFDM @ 28 GHz	SPW DSIC	- / -	25 / 60	3 / 7	18 dB	373	563
EIRP	43 dBm	SPH DSIC	3 / 7	25 / 60	- / -	25 dB	373	424
Received power ($d[n]$)	11 dBm	SPHW DSIC	3 / 7	25 / 60	3 / 7	29 dB	402	645

that also the LNA is distorting the SI waveform. All in all, the obtained results, together with the complexity analysis, indicated a very favorable performance-complexity trade-off of the proposed spline-based Hammerstein-Wiener model.

ACKNOWLEDGMENT

This work was financially supported by the Academy of Finland under the projects 301820, 323461 and 319994.

REFERENCES

[1] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," in *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4. ACM, 2013, pp. 375–386.

[2] M. Chung, M. S. Sim, J. Kim, D. K. Kim, and C. b. Chae, "Prototyping real-time full duplex radios," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 56–63, Sep. 2015.

[3] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Transactions on Wireless Communications*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[4] D. Korpi, J. Tamminen, M. Turunen, T. Huusari, Y.-S. Choi, L. Anttila, S. Talwar, and M. Valkama, "Full-duplex mobile device: Pushing the limits," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 80–87, Sep. 2016.

[5] D. Korpi, Y. Choi, T. Huusari, L. Anttila, S. Talwar, and M. Valkama, "Adaptive nonlinear digital self-interference cancellation for mobile inband full-duplex radio: Algorithms and RF measurements," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–7.

[6] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[7] P. Pascual Campo, D. Korpi, L. Anttila, and M. Valkama, "Nonlinear digital cancellation in full-duplex devices using spline-based Hammerstein model," in *IEEE Global Communications Conference (GLOBECOM)*, Dec. 2018.

[8] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Nonlinear spline adaptive filtering," *Signal Processing*, vol. 93, no. 4, pp. 772–783, 2013.

[9] 3GPP Tech. Spec. 38.104, "NR; Base Station (BS) radio transmission and reception," v15.4.0 (Release 15), Dec. 2018.

[10] M. Frerking, *Digital signal processing in communications systems*. Springer Science & Business Media, 2013.

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

[12] C. De Boor, *A practical guide to splines*. Springer-Verlag New York, 1978.

[13] H. Prautzsch, W. Boehm, and M. Paluszny, *Bézier and B-spline techniques*. Springer Science & Business Media, 2013.

[14] P. P. Campo, V. Lampu, L. Anttila, A. Brihuega, M. Allen, Y. Guo, and M. Valkama, "Closed-loop sign algorithms for low-complexity digital predistortion: Methods and performance," *IEEE Transactions on Microwave Theory and Techniques*, 2020.

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

PUBLICATION 8

Efficient inverse covariance matrix estimation for low-complexity closed-loop DPD systems

P. Pascual Campo, L. Anttila, V. Lampu, Y. Guo, N. Wang and M. Valkama

*2021 IEEE MTT-S International Wireless Symposium (IWS)*2021, 1–3
DOI: 10.1109/IWS52775.2021.9499480

Publication reprinted with the permission of the copyright holders

Efficient Inverse Covariance Matrix Estimation for Low-Complexity Closed-Loop DPD Systems

Pablo Pascual Campo[1], Lauri Anttila[1], Vesa Lampu[1], Yan Guo[2], Neng Wang[2], and Mikko Valkama[1]
[1]Department of Electrical Engineering, Tampere University, Finland
[2]Wireless Terminal Algorithm Development Department, HiSilicon Technologies Company

***Abstract*—This paper studies closed-loop digital predistortion systems, with special focus on linearization of mmW active antenna arrays. Considering the beam-dependent nonlinear distortion and very high DPD processing rates, a modified self-orthogonalized (SO) learning solution is proposed, which is capable of reducing the computational complexity compared to other similar solutions, while at the same time obtaining a comparable linearization performance. The modified SO consists of a novel method for efficiently calculating the inverse of the input data covariance matrix. Thorough RF measurement results at 28 GHz band featuring a state-of-the-art 64 element active array and channel bandwidths up to 800 MHz, are reported. A complexity analysis is also carried out which, together with the obtained results, allow to asses the performance-complexity trade-offs. Altogether, the results show that the proposed methods can facilitate efficient mmW active antenna array linearization.**

***Index Terms*—Array transmitters, mmW frequencies, nonlinear distortion, digital predistortion, self-orthogonalization, covariance matrix, real-time complexity, EVM, TRP ACLR.**

I. Introduction

Modern communication systems utilize spectrally efficient waveforms which typically infer high peak-to-average power ratio (PAPR) values, complicating the power-efficient operation of power amplifiers (PAs). In order to ensure a high efficiency in the transmitter system while still maintaining low levels of distortion, digital predistortion (DPD) can be applied. Various DPD techniques have been studied in the literature, with good examples shown in [1], [2], and references therein.

One particular modern DPD use case is the linearization of mmW antenna arrays [3], [4]. Linearizing such frequency range 2 (FR-2) systems is generally challenging, as the effective nonlinear distortion has been reported to be dependent on the array beam direction [4], [5]. This issue calls for fast DPD tracking to estimate the DPD as the beam is steered. Also, the 3GPP 5G NR Release 15 [6] already considers very wide channel bandwidths (BWs), which lead to high DPD processing rates. These issues call for low-complexity DPD solutions, which is the main focus of this paper.

Good overviews of 5G mmW array transmitters can be found in [3], [4], while [7]–[10] focus more deeply on DPD and linearization methods. In [7], authors proposed an efficient memory polynomial (MP) DPD with decorrelation based learning. However, an additional basis function (BF) orthogonalization was applied, increasing the processing complexity of the model. In [8], authors proposed a closed-loop 1-bit observation system in combination with a sign-based GN algorithm. In [9], a closed-loop MP model was presented, where the DPD model was estimated with damped Gauss-Newton (GN), in combination with a sign regressor algorithm (SRA). The SRA, however, can lead to rank-deficiencies in the BF matrix, while extra Walsh-Hadamard matrix transformations are also needed. Finally, [10] presented an MP model where the polynomials were replaced with look-up tables (LUTs), reducing the overall complexity. However, the model performance was also degraded with respect to canonical MP model, due to the quantization effects of the limited-size LUTs.

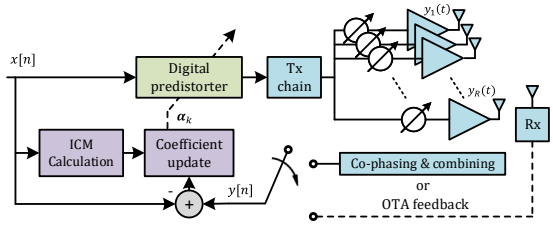


Fig. 1. Closed-loop DPD system for mmW active array linearization. The main novelty in this paper is in the DPD learning system.

In this article, in the context of active array transmitters shown in Fig. 1, a closed-loop MP DPD system in combination with a novel reduced-complexity self-orthogonalized (SO) learning rule is adopted. Specifically, an efficient approach to estimate the inverse covariance matrix (ICM) is presented, which constitutes the most complex term in the SO rule. This is particularly so in dynamic scheduling based systems where the waveform is also dynamic, in terms of e.g. modulation order and allocated BW. The modified learning rule achieves great complexity reductions in the learning path, while at the same time obtaining a similar linearization performance when compared to the original SO or reference GN methods. The adopted closed-loop DPD system together with the inherent low complexity of the novel SO learning algorithm, allow for continuous learning and fast real-time adaptation. Finally, to assess the performance of the proposed techniques, RF measurements at 28 GHz mmW band are carried out, utilizing a state-of-the-art 64 element active antenna array and channel BWs up to 800 MHz. The obtained results, along with a complexity analysis, show that efficient linearization can be obtained through the proposed methods, meeting in all cases the 3GPP NR Release 15 [6] requirements at FR-2.

II. DPD System and Proposed Method

The proposed method builds on the classical MP DPD model [1], whose input vector $\psi[n]$ at time instant n, with polynomial order P and memory depth M, is shown in (1). It

$$\psi[n] = \Big(x[n] \ \cdots \ x[n-M+1] \ x[n]|x[n]|^2 \ \cdots x[n-M+1]|x[n-M+1]|^2 \cdots \ x[n]|x[n]|^{P-1} \cdots \ x[n-M+1]|x[n-M+1]|^{P-1}\Big)^T \quad (1)$$

is combined with the closed-loop SO learning rule to estimate the DPD coefficients, $\boldsymbol{\alpha}_k$, with low complexity. The SO reads

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{s}}(\mathbf{R}^*)^{-1}\boldsymbol{\Omega}_k^H \mathbf{e}_k, \quad (2)$$

where k indicates the iteration index, μ_{s} is the learning rate, $\mathbf{e}_k$ is the closed-loop error signal, $\boldsymbol{\Omega}_k$ is the BF matrix stacking $\psi[n]$ for different time instants, and $\mathbf{R} = \mathrm{E}[\psi[n]\psi^H[n]]$ is the ensemble covariance matrix (CM). Using the classical sample estimation method, the CM can in practice be estimated as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \psi[n]\psi^H[n], \quad (3)$$

where N is the total number of input vectors over which the estimate is calculated. The proposed method, instead of directly calculating (3), utilizes the Bussgang's theorem to estimate the CM. Thus, a Gaussian input signal is assumed in the following derivations, while the practical experiments are carried out with true OFDM-based waveforms.

The first step is to express the CM exclusively as a function of the autocorrelation function of the input signal, denoted herein as $x[n]$. The autocorrelation function is defined as

$$R_{\mathrm{X}}(\tau) = \mathrm{E}\{x[n]x^*[n-\tau]\}, \quad (4)$$

where τ is a sample delay. The second-order terms in $\mathbf{R}$, on its first line, are obtained directly from (4), as $\mathbf{R}(1,1){=}R_{\mathrm{X}}(\tau{=}0)$, $\mathbf{R}(1,2){=}R_{\mathrm{X}}(\tau{=}1)$, and, in general, $\mathbf{R}(1,M){=}R_{\mathrm{X}}(\tau{=}M)$. Higher order terms in the CM (e.g., $\mathrm{E}\{x[n]|x[n-\tau]|^p\}$, $p = 2,4,\dots$), stemming from the nonlinear BFs, can be expressed through the cross-correlation function, as

$$R_{\mathrm{XY}}(\tau) = \mathrm{E}\{x[n]y^*[n-\tau]\}, \quad (5)$$

where $y[n] = f(x[n])$ is a nonlinear function of $x[n]$. To this end, the Bussgang's theorem states that the cross-correlation of a Gaussian signal passing through a nonlinear operator can be expressed as the product between its autocorrelation and a scaling constant. Formally, this is expressed as

$$R_{\mathrm{XY}}(\tau) = \xi R_{\mathrm{X}}(\tau), \quad (6)$$

where the Bussgang's coefficient, ξ, can be obtained assuming complex-circular Gaussian distribution as

$$\xi = \frac{1}{\pi\sigma_x^4}\int_{-\infty}^{\infty} x^*[n] f(x[n]) e^{-\frac{|x[n]|^2}{\sigma_x^2}} dx. \quad (7)$$

In (10), σ_x^2 denotes the variance of $x[n]$, which is typically smaller than 1 to obtain a small condition number in the CM. Using (6), the higher-order CM terms can be expressed as

$$\mathrm{E}\{x[n]x^*[n]|x[n]|^2\} = \xi_1 R_{\mathrm{X}}(\tau=0), \quad (8)$$
$$\mathrm{E}\{x[n]x^*[n-1]|x[n-1]|^2\} = \xi_1 R_{\mathrm{X}}(\tau=1), \quad (9)$$
$$\mathrm{E}\{x[n]x^*[n]|x[n]|^4\} = \xi_2 R_{\mathrm{X}}(\tau=0), \quad (10)$$
$$\mathrm{E}\{x[n]x^*[n-1]|x[n-1]|^4\} = \xi_2 R_{\mathrm{X}}(\tau=1), \quad (11)$$
$$\vdots$$

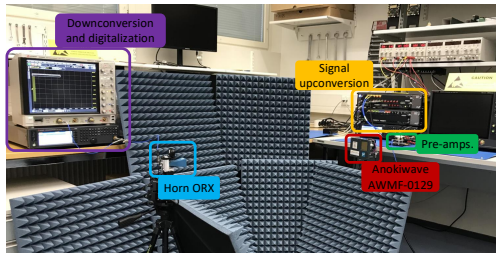


Fig. 2. The 5G mmW/FR-2 OTA measurement setup utilized to carry out the experimental measurements at 28 GHz.

where ξ_{x} is the corresponding Bussgang's coefficient. By having all terms in the CM expressed as a function of $R_{\mathrm{X}}(\tau)$ and the Bussgang's coefficients, the complete CM reads, for a generic polynomial order P and memory depth M, as

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \mathbf{R}_2 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} \\ \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{R}_3 & \cdots & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\lfloor \frac{P}{2} \rfloor} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+1} & \mathbf{R}_{\lfloor \frac{P}{2} \rfloor+2} & \cdots & \mathbf{R}_{2\lfloor \frac{P}{2} \rfloor} \end{bmatrix}, \quad (12)$$

where each subindex indicates the corresponding Bussgang's coefficient. Additionally, the sub-matrix $\mathbf{R}_k \in \mathbb{C}^{M\times M}$ is an Hermitian Toeplitz matrix, defined by the vector $\mathbf{v}_k = \xi_k[R_{\mathrm{X}}(\tau=0) \ R_{\mathrm{X}}(\tau=1) \ \cdots \ R_{\mathrm{X}}(\tau=M)]^T$, where $\xi_0 = 1$. Note that only the submatrices appearing in the first row and last column of (12) need to be calculated to build the whole CM for a given signal. The expression in (12) can be alternatively expressed as a Kronecker product, as

$$\mathbf{R} = \Xi \otimes \mathbf{R}_0, \quad (13)$$

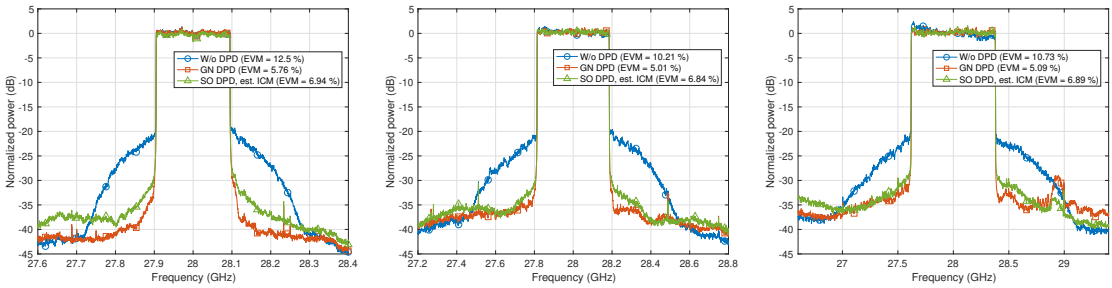
where Ξ contains the set of Bussgang's coefficients. Then, the ICM is directly obtained as

$$\mathbf{R}^{-1} = \Xi^{-1} \otimes \mathbf{R}_0^{-1}. \quad (14)$$

Thus, to calculate the ICM, only the autocorrelation function of $x[n]$ and the Bussgang coefficients are needed, both of which can be calculated in advance. The Kronecker formulation further reduces the required computational complexity.

III. RF Measurement Results

1) Measurement Setup: All the experiments are carried out with the state-of-the-art setup depicted in Fig 2. It consists of a Keysight M8195A to provide the I/Q data samples, additional equipment to upconvert the signal to 28 GHz, a pre-amplification stage, and finally the Anokiwave AWMF-0129 active antenna array. The signal is measured over-the-air (OTA), at an effective isotropic radiated power (EIRP) of +40.5 dBm, while then downconverted, digitized, and taken to a host PC to execute the algorithms. In all cases, the closed-loop MP DPD is configured with $P = 9$, $M = 4$, and $K = 20$



(a) (b) (c)

Fig. 3. OTA linearization performance at EIRP $\approx$ +40.5 dBm with the proposed SO learning rule (SO DPD, est. ICM), in (a) 200 MHz, (b) 400 MHz, and (c) 800 MHz channel bandwidth cases. The GN method is also implemented, measured and shown for reference.

TABLE I
LINEARIZATION PERFORMANCE AND AVERAGE LEARNING COMPLEXITY PER SAMPLE IN THE 800 MHZ CHANNEL BW CASE.

	EVM (%)	**TRP ACLR (dBc)**	**Learning path complexity (real mul. per lin. sample)**
Classical SO DPD	6.02	31.2	186.9
Proposed SO DPD	6.89	30.4	81.42
Reference GN DPD	5.09	32.3	1,681

ksamples per iteration, with a total of 20 DPD iterations. The results also include and show the GN learning solution [8], [9] as a reference, which reads

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \mu_{\mathrm{g}} \left(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k \right)^{-1} \boldsymbol{\Omega}_k^H \mathbf{e}_k, \tag{15}$$

where μ_{g} is the learning rate. The performance of this model is generally better compared to SO, since the term $\left(\boldsymbol{\Omega}_k^H \boldsymbol{\Omega}_k \right)^{-1}$ is calculated in all DPD iterations, however, it also clearly infers larger complexity.

The adopted signals are 5G NR Rel-15 compliant OFDM waveforms, with subcarrier spacing of 120 kHz, covering BWs of 200 and 400 MHz at FR-2. An experiment with a further extended BW of 800 MHz is also included, with the aim of pushing the performance boundaries of the DPD system.

2) Measurement Results: The measured PSDs are depicted in Fig. 3. In all BW cases, the proposed SO learning solution provides a very favorable linearization performance, very close to the GN solution. The error vector magnitude (EVM), adjacent channel leakage ratio (ACLR), and complexity numbers are shown in Table I, for the widest signal BW of 800 MHz, where the ACLR is measured using the total radiated power (TRP) approach. The 5G NR EVM limit of 8%, and the TRP ACLR limit of 28 dBc [6] are fulfilled in all cases, regardless of the 56% complexity reduction with respect to classical SO and 95% complexity reduction with respect to GN. These performance and complexity results demonstrate that the proposed solution is an intriguing approach for the efficient linearization of mmW active antenna arrays.

IV. CONCLUSIONS

In this paper, an efficient method to estimate the inverse covariance matrix of the DPD basis function samples was proposed. The proposed method was shown to provide large complexity reductions in the context of self-orthogonalized (SO) DPD learning applications. The complexity-performance trade-offs of the proposed overall DPD system were demonstrated through extensive RF measurement results at 28 GHz, featuring a state-of-the-art Anokiwave AWMF-0129 active array, and very wide channel BWs up to 800 MHz. The obtained performance results, together with the complexity analysis, indicate that efficient mmW active antenna array linearization can be achieved through the proposed technique.

ACKNOWLEDGMENT

This work was supported in part by Huawei Technologies, and in part by the Academy of Finland under the projects 301820, 323461, 332361 and 319994.

REFERENCES

[1] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, 2009.
[2] P. Pascual Campo *et al.*, "Gradient-adaptive spline-interpolated LUT methods for low-complexity digital predistortion," *IEEE Trans. Circuits Syst. I*, pp. 1–14, 2020.
[3] X. Wang, Y. Li, C. Yu, W. Hong, and A. Zhu, "Digital predistortion of 5G massive MIMO wireless transmitters based on indirect identification of power amplifier behavior with OTA tests," *IEEE Trans. Microw. Theory Tech.*, vol. 68, no. 1, pp. 316–328, 2020.
[4] C. Fager *et al.*, "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.
[5] P. Pascual Campo *et al.*, "Closed-loop sign algorithms for low-complexity digital predistortion: Methods and performance," *IEEE Trans. Microw. Theory Tech.*, 2020.
[6] 3GPP Tech. Spec. 38.141-2, "NR; Base Station (BS) conformance testing, Part 2," v15.1.0 (Release 15), March 2019.
[7] M. Abdelaziz, L. Anttila, A. Kiayani, and M. Valkama, "Decorrelation-based concurrent digital predistortion with a single feedback path," *IEEE Trans. Microw. Theory Tech.*, vol. 66, no. 1, pp. 280–293, 2018.
[8] H. Wang, G. Li, C. Zhou, W. Tao, F. Liu, and A. Zhu, "1-bit observation for direct-learning-based digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 7, pp. 2465–2475, 2017.
[9] N. Guan, N. Wu, and H. Wang, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 921–923, Oct 2018.
[10] A. Molina, K. Rajamani, and K. Azadet, "Digital predistortion using lookup tables with linear interpolation and extrapolation: Direct least squares coefficient adaptation," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 3, pp. 980–987, 2017.