Behavioral model on transmitter power amplifier and IQ imbalance

Fu, Kai

2014

Fu, K. (2014). Behavioral model on transmitter power amplifier and IQ imbalance. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/61666>

<https://doi.org/10.32657/10356/61666>

Behavioral Model on Transmitter Power Amplifier and IQ Imbalance

Fu Kai

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2014

Acknowledgements

I am very grateful to my supervisor, Assoc. Prof. Law Choi Look, for the invaluable guidance that he has given to me throughout my research, the spirits of diligence and strictness that he has shown to me in our every discussion, and the kindness and patience that he has awarded to me on my every mistake. There were countless times that Prof. Law directed and helped me solve the problems when I was lost and felt nowhere to go. I am so indebted to Prof. Law for the spirits that I have learnt from him and that will certainly and greatly benefit my future life.

Without the love, encouragement, and sacrifice from my dear wife, Lili, I could achieve nothing but be ruined in a total failure. She is the source of my happiness, the fountain of my strength, and the base of my belief. For her ever encouraging smile, I owe my every achievement to my wife.

I am in great debt to my parents, not only for their once-a-week spiritual support in the call from thousands of miles away, but for their never dwindling care, understanding, and love. I am in great debt to my mother-in-law, who has not only taught me how to face the difficulties, but how to live a meaningful life. And I am in great debt to my dear son and daughter, Bowen and Jiaying, from whom I obtain so much happiness, of whom I feel so proud, and upon whom I regret so much to have spent so little time.

The help from all my friends and my colleagues in Infinitus Center is indispensable and invaluable; the support from all my relatives, friends, and classmates is crucial and priceless.

To my wife, I dedicate this thesis.

Table of Contents

Acknowledgements	i
Summary	v
List of Figures	viii
List of Tables	xiv
List of Abbreviations	xv
List of Principle Symbols	xix
Chapter 1 Introduction.....	1
1.1 Background Information.....	1
1.2 Motivation and Objectives	3
1.3 Major Contribution.....	6
1.4 Arrangement of the Thesis	8
Chapter 2 Literature Review	10
2.1 TX IQ Imbalance and PA Distortions	10
2.2 Behavioral Models of Transmitter Distortions	15
2.3 Behavioral Models of Power Amplifier.....	18
2.4 TX IQ Imbalance and Behavioral Models.....	31
2.5 Current Results on Unified Model of PA + IQ Imbalance	38
2.6 Conclusion	39
Chapter 3 Test Bed for TX IQ Imbalance and Power Amplifier Behavioral Characterization and Modeling	40
3.1 Introduction.....	40

3.2	Proposed PA Test Bed and Signal Processing	43
3.3	TX IQ Imbalance Test Bed	60
3.4	Joint TX IQ Imbalance and PA Test Bed	61
3.5	Performance Metrics Design	62
3.6	Hardware Design Consideration	66
3.7	Performance Assessment	69
3.8	Conclusion	80
Chapter 4 A New Power Amplifier Predistortion Architecture Based on Memory Polynomial Model		82
4.1	Introduction	82
4.2	Memory Effects and Memory Polynomial Model	83
4.3	New DPD Architecture Based on MPM	86
4.4	Simulation Results	93
4.5	Conclusion	96
Chapter 5 Frequency-Domain Predistortion of TX IQ Imbalance of Significant Memory Effects for OFDM Systems		98
5.1	Introduction	98
5.2	Signal and System Model	101
5.3	Drawback of Time-Domain Predistortion	110
5.4	Frequency-Domain Predistortion on TX IQ Imbalance of Long Memory Effects	114
5.5	Predistorter Identification and Sensitivity to Noise	116
5.6	Simulation Results	125
5.7	Experiment Results	132

5.8	Conclusion	135
Chapter 6	Novel Neural Network Model of Power Amplifier + IQ Imbalance	137
6.1	Introduction.....	137
6.2	Unified Model.....	138
6.3	Limitation on Polynomial Based Models	139
6.4	Proposed Neural Network Model	143
6.5	Experiment Results.....	155
6.6	Conclusion	176
Chapter 7	Conclusion and Future Works	177
7.1	Conclusion	177
7.2	Future Works	180
Author's Publications.....		182
Bibliography		183
Appendix A	Least Squares	193
A.1	Traditional LS problem.....	193
A.2	Total Least Squares	193
A.3	Data Least Squares	194
Appendix B	Singular Value Decomposition	197
Appendix C	Toeplitz and Circulant Matrices	201

Summary

This thesis is dedicated to the subject of the behavioral model of two most important analog impairments in modern wireless transmitters – in-phase/quadrature (IQ) imbalance distortion and power amplifier (PA) nonlinear memory distortion. Despite their distinct physical characteristics, and unlike conventional ways of treating them differently and separately, this thesis is to treat these two seemingly different distortions equally and model them uniformly. Specifically, the main research efforts have gone through the following two stages: In the first stage, these two distortions are treated equally while independently, and compensated using the same methodology – digital predistortion. In this stage, the main objective is to find a better behavioral model for each distortion. Then in the next stage, these two distortions are merged as one black box, and characterized by a single behavioral model. The objective of this stage is to unify the view of the two distortions, abstract it in the implementation, and save the resources by unifying and simplifying the model.

After the introduction chapter, the background of the two analog distortions – transmitter (TX) in-phase/quadrature imbalance and power amplifier – are presented, followed by the general discussion of system-level simulation and digital predistortion, where the quality of the behavioral model determines both performances. Then a comprehensive literature review is engaged on the subjects of transmitter in-phase/quadrature Imbalance and power amplifier, from where it is clear that the current available behavioral models cannot satisfy the demanding requirements, and hence efforts are necessary to find a better, simpler, more accurate, while free-of-side-effects model.

As the prerequisites of behavioral characterization, the issue of test bed is researched and developed. A competent test bed should be accurate, sufficient, and flexible. Specifically, it should provide sufficient and flexible facilities to engage and complete the processes of radio frequency (RF) signal envelope recovery, behavior characterization, model identification, model test and verification. In this thesis, a test bed based on an oscilloscope or other general purpose data acquisition systems, which works as analog to digital converter (ADC) with a proper radio frequency bandwidth and sampling rate, is proposed. The common impairments, e.g. transmitter in-phase/quadrature imbalance, channel delay, frequency offset, and carrier phase offset, are all well compensated. The accurately recovered envelopes of the power amplifier's input and output signals are used for power amplifier behavioral characterization and modeling. The experiment results show a very accurate radio frequency signal envelope recovery, and a good performance of power amplifier behavioral modeling.

In the next two chapters, the first research stage is carried on: Find a better behavioral model for each distortion. Firstly, a predistortion scheme to compensate the power amplifier nonlinear memory distortion is proposed. Unlike the classical memory polynomial model (MPM) that predistorts the amplitude and phase of the signals simultaneously, this thesis describes a new memory polynomial model based predistortion architecture that predistorts amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) separately. The idea comes from the inherent physical characteristics of the power amplifier – the amplitude-to-amplitude and amplitude-to-phase relationships. The new architecture can achieve a significantly better performance than the classical memory polynomial model architecture.

Subsequently, a new simple frequency-domain transmitter predistorter is proposed to compensate transmitter in-phase/quadrature imbalance. Frequency-domain predistortion can perfectly compensate long memory effects, without introducing inter-symbol interference (ISI) for orthogonal frequency-division multiplexing (OFDM) signal. And the numerical issue is carefully analyzed, and some robust numerical methods are proposed to achieve high accuracy. The complexity of these methods is also analyzed.

For the second research stage of merging these two distortions into a single module, a new unified behavioral model for both transmitter in-phase/quadrature imbalance distortion and power amplifier distortion is proposed. Traditionally, these two distortions are separately modeled. In this thesis, the behavior of the two distortions are unified, and characterized by a single model. Rectangular Structured Focused Time-Delay Neural Network (RSFTDNN) is proposed to uniformly model transmitter in-phase/quadrature imbalance and power amplifier distortions. As a result, the physical distortions in the analog circuits are further abstracted. It also saves the computation resources. Unlike the polynomial based model, which suffers from inaccuracy for deeply nonlinear system, the proposed Rectangular Structured Focused Time-Delay Neural Network shows a strong ability and high accuracy. Two cases of real experiments are carried out, where Rectangular Structured Focused Time-Delay Neural Network model shows a much better performance than the polynomial based model in the sense of the model accuracy.

List of Figures

Fig. 2-1 AM/AM relation of power amplifier	12
Fig. 2-2 AM/PM relation of power amplifier	12
Fig. 2-3 Overall diagram of a wireless transmitter, where the signal with the subscript p is the RF passband signal with carrier frequency, while the signal with the subscript l is the corresponding low-pass equivalent signal. The index n means the digital signal, while index t means the continuous analog signal after DAC. And θ_I and θ_Q are the phase imbalance owing to LO imbalance.	14
Fig. 2-4 Prototype of digital predistortion for analog impairments, where the discrete low-pass equivalent signals are used to represent the analog RF signals	17
Fig. 2-5 A typical Time-Delay Neural Network (TDNN) with 2 inner layers	28
Fig. 2-6 A typical Recurrent neural network (RNN) with 2 inner layers and single output. The input signal is delayed at the input layer, where z^{-1} means one sample delay. $f(\cdot)$ is the activation function.	29
Fig. 2-7 RBFNN with $L + 1$ input nodes, M hidden nodes, and two output nodes. The two output nodes correspond to the AM/AM and AM/PM distortion, respectively [67].	30
Fig. 2-8 Classification of IQ imbalance compensation methods: (a) pre-compensation for TX IQ imbalance. (b) post-compensation for RX IQ imbalance. (c) post-compensation for joint TX and RX IQ imbalance.	31
Fig. 2-9 Prototype of digital predistortion of TX IQ imbalance	35
Fig. 2-10 TX IQ Imbalance Predistortion	35
Fig. 2-11 Behavioral model of the TX IQ imbalance distortion [117]. And $\alpha_{II}, \alpha_{IQ}, \alpha_{QI}, \alpha_{QQ}$ are the weights for each path.	36
Fig. 3-1 Overall Structure of Test Bed for PA behavioral modeling	44
Fig. 3-2 Block diagram of digital signal processing	45

Fig. 3-3 Linear predistorter to compensate IQ imbalance.....	47
Fig. 3-4 IQ imbalance model and linear predistorter	47
Fig. 3-5 Digital demodulation	50
Fig. 3-6 General block diagram of down sampler.....	51
Fig. 3-7 Direct-Form FIR filter structure with embedded down sampler	52
Fig. 3-8 Polyphase filter structure	53
Fig. 3-9 Signal structure	54
Fig. 3-10 Final digital demodulation with LPFs, down samplers, and the first step of time alignment – finding optimal down sampling position.....	57
Fig. 3-11 Transmitter digital signal processing for TX IQ imbalance Test Bed.....	60
Fig. 3-12 Test Bed for TX IQ imbalance. The dotted block represents an additional IQ imbalance that can be manually added in the transmitter.....	61
Fig. 3-13 Test Bed for joint TX IQ Imbalance and PA.....	62
Fig. 3-14 Welch method: segmentation with 50% overlap	65
Fig. 3-15 Down Converter Channels.....	67
Fig. 3-16 Calibration of the nonlinearity of down-converters: Gain versus input power level for Channel A and Channel C.	68
Fig. 3-17 ZVE-8G gain (dB) versus frequency	72
Fig. 3-18 Output 1dB compression point power versus frequency	72
Fig. 3-19 Constellation: 16QAM+BPSK, without compensating IQ imbalance in modulator, EVM is -34.3 dB.	74
Fig. 3-20 Constellation: 16 QAM+BPSK, with compensated IQ imbalance in modulator, EVM is -42.3 dB.	74
Fig. 3-21 Time-domain waveforms comparison: Removing DUT, the real part of recovered envelope waveform by Channel C versus the real part of transmitted baseband waveform.	78

Fig. 3-22 Time-domain waveforms comparison: Removing DUT, the image part of recovered envelope waveform by Channel C versus the image part of transmitted baseband waveform.....	78
Fig. 3-23 Recovered AM/AM behavior of ZVE-8G HPA: The curve is drawn from the recovered envelopes of PA's input and output signals.....	79
Fig. 3-24 Recovered AM/PM behavior of ZVE-8G HPA: The curve is drawn from the recovered envelopes of PA's input and output signals.	79
Fig. 4-1 Linearized PA block diagram.....	92
Fig. 4-2 New DPD architecture using MPM.....	93
Fig. 4-3 Normalized PSD of the signals, using memory polynomial model predistorter. (a) Predistorted by the traditional architecture. (b) Predistorted by the new architecture. (c) Original input signal	95
Fig. 4-4 ACPR over many frames. (a) Left ACPR with the traditional architecture. (b) Left ACPR with the new architecture. (c) Right ACPR with the traditional architecture. (d) Right ACPR with the new architecture.....	95
Fig. 4-5 constellation of the signals at the receiver after equalization. (a) Traditional architecture. (b) New architecture	96
Fig. 4-6 EVM comparison over time at the receiver.....	97
Fig. 5-1 Block Diagram of TX IQ Imbalance: $\tilde{h}_I(t)$ and $\tilde{h}_Q(t)$ are the real-valued filters modeling the IQ imbalance of memory effects due to the imperfection of the DACs and the LPFs, and $a_I, \theta_I, a_Q, \theta_Q$ are the gain and phase imbalance owing to LO imbalance	101
Fig. 5-2 Time-domain TX IQ imbalance predistortion, where “Prd” stands for “predistortion”.....	111
Fig. 5-3 Time-domain predistorter	112
Fig. 5-4 Proposed frequency-domain long-memory TX IQ imbalance predistortion	115

Fig. 5-5 Behavioral model of the TX IQ imbalance distortion in the simulation	126
Fig. 5-6 EVM performance versus SNR values, under 64 QAM scheme.....	128
Fig. 5-7 EVM performance versus memory length of the TX IQ imbalance distortion under 64 QAM scheme	129
Fig. 5-8 Comparison of the EVM performance over each subcarrier	131
Fig. 5-9 Condition number comparison between the proposed frequency-domain predistorter and the time-domain predistorter.	131
Fig. 5-10 Comparison with other frequency-domain TX IQ imbalance predistortion methods. Different lengths of training sequence are swept, where the length of training sequence is in the number of OFDM symbols.....	132
Fig. 5-11 Experiment setup for frequency domain TX IQ imbalance predistortion...	133
Fig. 5-12 Constellation comparison in real experiments: The left figure is the 64-QAM constellation suffering from TX IQ imbalance of long-memory; the right figure is the corresponding constellation after using the proposed frequency-domain predistortion to compensate.	135
Fig. 6-1 IQ imbalance and PA distortions at the wireless transmitter. The index n means the digital signal, while index t means the continuous analog signal after DAC.	144
Fig. 6-2 Diagram of TX IQ imbalance and PA impairments.....	144
Fig. 6-3 First form of unified model for joint TX IQ imbalance and PA impairments	146
Fig. 6-4 Second form of unified model for joint TX IQ imbalance and PA impairments	147
Fig. 6-5 Linear Gain System	147
Fig. 6-6 Rectangular structure of the second form of unified model	149

Fig. 6-7 RSFTDNN model, where z^{-1} means one sample delay. The number of inner layers, and the number of neurons of each inner layer can be varying according to different devices and requirements.....	151
Fig. 6-8 Design of neuron: the i^{th} neuron in the $l+1^{th}$ layer.....	152
Fig. 6-9 Experiment devices setup diagram	156
Fig. 6-10 Learning curve of RSFTDNN model	159
Fig. 6-11 Time domain waveform comparison	160
Fig. 6-12 AM/AM and AM/PM relations comparison	160
Fig. 6-13 Spectrum comparison, including the spectrum of the error between the measured and modeled results.....	161
Fig. 6-14 Accuracy comparison between RSFTDNN model and MPM-Dual-Input model, and the condition number of MPM-Dual-Input model.	162
Fig. 6-15 Resources consumption comparison between RSFTDNN model and MPM-Dual-Input model: memory space consumption and simulation time consumption. The comparison is based on a Hewlett-Packard Z210 workstation with Intel Xeon CPU E31225 @3.10GHz and 4.00 GB RAM memory.....	163
Fig. 6-16 Additional IQ imbalance added in baseband	164
Fig. 6-17 Experiment devices setup diagram with additional TX IQ imbalance	164
Fig. 6-18 Learning curve of RSFTDNN model	166
Fig. 6-19 Time domain waveform comparison	167
Fig. 6-20 AM/AM and AM/PM relations comparison	168
Fig. 6-21 Spectrum comparison, including the spectrum of the error between the measured and modeled results.....	168
Fig. 6-22 Accuracy comparison between RSFTDNN model and MPM-Dual-Input model, and the condition number of MPM-Dual-Input model	169
Fig. 6-23 Resources consumption comparison between RSFTDNN model and MPM-Dual-Input model: memory space consumption and simulation time	

consumption. The comparison is done based on the same computer as the previous experiment.	169
Fig. 6-24 Time domain waveform comparison for the case of Section 6.5.2, using OFDM testing signal	170
Fig. 6-25 Time domain waveform comparison for the case of Section 6.5.3, using OFDM testing signal	171
Fig. 6-26 AM/AM and AM/PM relation comparison for the case of Section 6.5.2, using OFDM testing signal	171
Fig. 6-27 AM/AM and AM/PM relation comparison for the case of Section 6.5.3, using OFDM testing signal	172
Fig. 6-28 Spectrum comparison for the case of Section 6.5.2, using OFDM testing signal, including the spectrum of the error between the measured and modeled results	172
Fig. 6-29 Spectrum comparison for the case of Section 6.5.3, using OFDM testing signal, including the spectrum of the error between the measured and modeled results	173
Fig. 6-30 Comparison of different layouts: number of inner layers, and number of neurons of each inner layer. It is for the case of Section 6.5.3, using Gaussian testing signal.....	175
Fig. 6-31 Comparison of different layouts: number of inner layers, and number of neurons of each inner layer. It is for the case of Section 6.5.3, using OFDM testing signal.....	175

List of Tables

Table 3-1 ZVE-8G specifications.....	71
Table 3-2 ZVE-8G datasheet.....	71
Table 3-3 EVM measurement results for Channel A and Channel C.....	75
Table 3-4 Channel A: REE values versus baseband sampling rate.....	76
Table 3-5 Channel C: REE values versus baseband sampling rate.....	76
Table 3-6 Channel A: REE value versus time alignment resolution.....	77
Table 4-1 The number of coefficients for different models, where K denotes nonlinear order, M denotes memory depth.....	91
Table 5-1 Identification Complexity Comparison.....	124
Table 5-2 Comparison of the number of coefficients: J is the number of subcarriers of the OFDM signal, while M is the length of time-domain predistorter	125
Table 5-3 Coefficients of the IQ imbalance model in “EVM vs SNR” simulation...	127
Table 5-4 Coefficients of the additional IQ imbalance model in the experiment	134
Table 5-5 EVM performance of different schemes in experiment measurements	134
Table 6-1 Physical parameters comparison between real device and RSFTDNN model.....	159
Table 6-2 Coefficients of the IQ imbalance model	165
Table 6-3 Physical parameters comparison between real device and RSFTDNN model.....	167

List of Abbreviations

1-D	One Dimensional
2-D	Two Dimensional
3-D	Three Dimensional
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
AC	Alternating Current
ACPR	Adjacent Channel Power Ratio
ADC	Analog-to-Digital Converter
ADS	Advanced Design System
AGC	Automatic Gain Control
AM/AM	Amplitude Dependent Amplitude Distortion
AM/PM	Amplitude Dependent Phase Distortion
ANFIS	Adaptive Neural Fuzzy Inference System
ANN	Artificial Neural Network
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BJT	Bipolar Junction Transistor
CDMA	Code Division Multiple Access
CFO	Carrier Frequency Offset
CP	Cyclic Prefix
CRLB	Cramer-Rao lower bound
DAC	Digital-to-Analog Converter
DC	Direct Current

DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DIDO	Dual Input Dual Output
DISO	Dual Input Single Output
DLS	Data Least Squares
DPD	Digital Predistortion
DSP	Digital Signal Processing
DUT	Device Under Test
EVM	Error Vector Magnitude
FELMS	Fast Exact Least Mean Squares
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
GI	Guard Interval
GPIB	General Purpose Interface Bus
HBT	Heterojunction Bipolar Transistor
HPA	High Power Amplifier
ISI	Inter Symbol Interference
IC	Integrated Circuit
ICI	Inter Carrier Interference
IF	Intermediate Frequency
IFFT	Inverse Fast Fourier Transform
IIR	Infinite Impulse Response
I.I.D.	Independent and Identically Distributed
IM	Intermodulation
IM3	Third Order Intermodulation Distortion

IMD	Intermodulation Distortion
INMFI	Inter-Non-Mirror-Frequency Interference
I/Q (IQ)	in-phase/quadrature
IRR	Image Rejection Ratio
KLT	Karhunen-Loeve Transform
LMS	Least Mean Squares
LNA	Low Noise Amplifier
LO	Local Oscillator
LPF	Low Pass Filter
LRLS	Lattice Based Recursive Least Squares
LS	Least Squares
LTE	Long Term Evolution
LUT	look up table
MFI	Mirror Frequency Interference
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean-Square Error
MPM	Memory Polynomial Model
NLMS	Normalized Least Mean Squares
NN	Neural Network
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PAE	Power Added Efficiency
PAPR	Peak-to-Average Power Ratio
PSD	Power Spectral Density

PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QR-RLS	QR Decomposition Based Recursive Least Squares
RBFNN	Radical-Basis Function Neural Network
REE	Relative Envelope Error
RF	Radio Frequency
RLS	Recursive Least Squares
RSA	Real-time Spectrum Analyzer
RX	Receiver
SC	Single Carrier
SER	Symbol Error Rate
SISO	Single Input Single Output
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
SW-LMS	Sliding Window Least Mean Squares
TAS	Time Alignment Sequence
TLS	Total Least Squares
TX	Transmitter
VLSI	Very Large Scale Integration
VSA	Vector Signal Analyzer
WiMax	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
ZF	Zero Forcing

List of Principle Symbols

Mathematics and Matrix Representations:

$\mathbf{A}$	Matrix or matrix variable
$[\mathbf{A}]_{N \times M}$	Matrix of N rows and M columns
$\mathbf{b}$	Vector or vector variables
$\mathbf{A}_i$	Matrix indexed for some purpose
$\mathbf{A}_{ij}$ or $(\mathbf{A})_{ij}$	The $(i, j)th$ entry of the matrix $\mathbf{A}$
$\mathbf{A}^n$	the n^{th} power of a square matrix
$\mathbf{A}^{-1}$	The inverse matrix of the matrix $\mathbf{A}$
$\mathbf{A}^\dagger$	The pseudo inverse matrix of the matrix $\mathbf{A}$
$\mathbf{b}_i$	Vector indexed for some purpose
b_i	Scalar indexed for some purpose
b	Scalar or scalar variables
$x(n)$	Discrete time signal
$x(t)$	Continuous time signal
$\Re\{x\}$	Real part of a scalar
$\Re\{\mathbf{x}\}$	Real part of a vector
$\Re\{\mathbf{A}\}$	Real part of a matrix
$\Im\{x\}$	Imaginary part of a scalar
$\Im\{\mathbf{x}\}$	Imaginary part of a vector
$\Im\{\mathbf{A}\}$	Imaginary part of a matrix

$\det(\mathbf{A})$	Determinant of the matrix $\mathbf{A}$
$Tr(\mathbf{A})$	Trace of the matrix $\mathbf{A}$
$diag(\mathbf{b})$	Diagonal matrix of the vector $\mathbf{b}$, i.e. $(diag(\mathbf{A}))_{ij} = \delta_{ij}b_i$
$eig(\mathbf{A})$	Eigenvalues of the matrix $\mathbf{A}$
$\sup$	Supremum of a set
$ x $	absolute value or amplitude of scalar variable x
$\ \mathbf{A}\ $	Matrix norm (subscript if any denotes what norm)
$\ \mathbf{A}\ _2$	2-norm of matrix
$\ \mathbf{A}\ _F$	<i>Frobenius</i> norm of matrix
$\mathbf{A}^T$	Transposed matrix
$\mathbf{A}^*$	Complex conjugated matrix
$\mathbf{A}^H$	Transposed and complex conjugated matrix (Hermitian)
$\chi(\mathbf{A})$	Condition number of the matrix $\mathbf{A}$
$span(\mathbf{A})$	Space spanned by the matrix $\mathbf{A}$
$\mathbf{A} \circ \mathbf{B}$	Hadamard (elementwise) product
$s(n) \otimes d(n)$	Convolution operation between two signals
$\mathbf{O}$	The null matrix with zeros in all entries
$\mathbf{0}$	The null vector with zeros in all entries
$\mathbf{1}$	The all-one column vector
$\mathbf{I}$	The identity matrix
$\mathbf{P}$	Projection Matrix

$\text{diag}\{\mathbf{b}\}$	Diagonal matrix with $\mathbf{b}$ on its diagonal
$\mathbf{U}, \Sigma, \mathbf{V}$	SVD computation matrices
λ	Lagrange multiplier
$\{\lambda_i\}$	Eigenvalues
$\{\sigma_i\}$	Singular values
dy/dx	Derivatives
$\partial y / \partial x$	Partial derivatives
$\nabla \mathbf{y}$	Gradient
$\Delta b, \Delta \mathbf{b}, \Delta \mathbf{A}$	Increment scalar, vector and matrix
$E\{\mathbf{x}\}, \text{var}\{\mathbf{x}\}$	Expectancy and variance of vector $\mathbf{x}$
$\text{cov}\{x, y\}$	Covariance of variable x and y
$\arg\{x\}, \angle x, \sphericalangle x$	Phase of scalar variable x
$\prod_{i=1}^N x_i$	Multiplication of the set of $\{x_i\}_{1 \leq i \leq N}$
$f^{-1}(x)$	Inverse function

Physical Representations:

$s(t), s(n)$	Baseband transmitted signal
$r(t), r(n)$	Baseband received signal
$s_p(t), r_p(t)$	Passband (RF) signals
$s_l(t), r_l(t)$	Equivalent low-pass signals

$s_I(n), s_Q(n)$	I component and Q component of signal $s(n)$
$S(f), S(k)$	Frequency domain continuous and discrete signals
$\hat{S}(k)$	The conjugate mirror frequency component
$\mathbf{s}^f$	Frequency domain signal vector
$h, \mathbf{h}$	Model, filter or channel coefficients (element and vector)
$u, \mathbf{u}, \mathbf{U}$	Model input (element, vector and matrix)
$y, z, \mathbf{y}, \mathbf{z}, \mathbf{Y}, \mathbf{Z}$	Model output
$x, \mathbf{x}, \mathbf{X}$	It can serve as either model input or output (dependent on the context); at the same, it can also serve as the solution of an equation (dependent on the context).
i, j, p, q	general variable Indices
m	Memory or tap index
k	Nonlinear order index, discrete frequency index
n	Discrete time index
t, τ	Continuous time
$e, \mathbf{e}, \mathbf{E}$	Error
$v, \mathbf{v}$	Noise
f	Continuous frequency
f_s	Sampling Rate Frequency
μ	Step size in adaptive identification methods
$\Phi_{i,(l+1)}(\cdot)$	Activation function for the i th neuron in the $l+1$ th layer.
$\alpha_{ij,(l+1)}$	Weight from the j th neuron in the l th layer to the i th neuron in the $l+1$ th layer

$\beta_{i,(l+1)}$	Bias for the i th neuron in the $l + 1$ th layer, respectively.
G	(power) gain
θ, ϕ	phase

Chapter 1 Introduction

1.1 Background Information

Modern wireless communication systems suffer power amplifier (PA) and IQ imbalance distortions at the transmitter. On one hand, power amplifier, digital to analog converter (DAC), reconstruction filter and up-converter are indispensable in modern wireless communication systems. While on the other hand, these analog circuits result in transmitter (TX) IQ imbalance and PA distortions on the signal, due to their imperfect implementation. And although it is possible to design and manufacture analog devices as perfectly as possible to reduce these distortions, the corresponding cost is quite expensive. Hence owing to cost-area-consideration, TX IQ imbalance and PA distortions are usually unavoidable in modern wireless communication systems.

Although the research on TX IQ imbalance and PA has been on-going for quite a while, modern wireless communication systems are also evolving quickly with demand for more energy efficiency and higher data rates. The former demand usually requires that the PA works in deeper nonlinear mode, while the later demand often requires that the signal has wider bandwidth. As a result, the modern power amplifier demonstrates deeper and deeper nonlinearity and more and more memory effects, while the TX IQ imbalance distortion shows more and more significant frequency-dependent effects.

The deep nonlinearity of PA distortion quickly degrades the quality of modern wideband signals, e.g., Orthogonal Frequency Division Multiplexing (OFDM) signals [1]. OFDM technology has been widely used in many modern communication systems, e.g. Wireless Local Area Networks (WLAN) [2], Wireless Metropolitan Area Networks (WMAN) [3] and Long Term Evolution (LTE) [4]. The bandwidths of those systems vary in different standards, and typical bandwidths are 16.6 MHz in WLAN [5], and 20 MHz in WiMax [3] and LTE [6]. OFDM signals have the problem of high peak-to-average-power ratio (PAPR), and hence suffer great quality-degradation from nonlinear distortions. It brings a big challenge to the design of the system – while OFDM signals demand linear circuits to reduce bit-error-rate (BER), modern power amplifiers are working in deep nonlinear mode to achieve high energy efficiency.

TX IQ imbalance [7, 8] is represented as the gain imbalance between I and Q-paths, and the shift from 90° phase difference between the two local oscillators (LOs). The distortion introduces so called Mirror Frequency Interference (MFI) [9]. And with a time-variant channel, e.g. by Doppler effect [10], IQ Imbalance can result in Inter-Non-Mirror-Frequency Interference (INMFI) [9]. Both MFI and INMFI can severely distort the transmitted signal, and result in information mistranslation and loss.

Meanwhile, as the modern transmitting signal has wider bandwidth, memory phenomenon becomes more significant. And memory phenomenon makes the compensation more complicated and difficult – The distortion is generated not only by the simultaneous input signal but also by the past ones. Consequently, any compensation method should also include the previous inputs. Both TX IQ Imbalance and PA distortions have significant memory phenomena.

Hence without careful consideration or design, TX IQ Imbalance and PA distortions can largely degrade the quality of the transmitting signal, resulting in poor BER performance, and interference on other signals of the adjacent channels due to the out-of-band spectral re-growth effects. From the viewpoint of the behaviors of the distortions, this thesis is dedicated to research on these two distortions to cater to the requirements of the evolving modern wireless communication systems – deep nonlinearity and significant memory effects.

Note that historically, the memory phenomenon is usually called *memory effects* in the field of power amplifier, while the same thing is called *frequency-dependent* or *frequency-selective* in the field of IQ imbalance. In this thesis, these notations are united, and “*memory effects*” is used to refer to the memory phenomenon of the device.

1.2 Motivation and Objectives

Power amplifier distortion and TX IQ imbalance distortion are two major analog distortions at the wireless transmitter. Hence, the following tasks are of great importance and necessity for a modern communication system:

1. Analyze the characteristics and influence of the TX IQ Imbalance and PA distortions, and predict the performance of the signal impaired by these distortions.
2. Find an efficient and low-cost method to compensate these two distortions.

System-level simulation is a good way to accomplish the first task. In system-level simulation, the signal is passed through the behavioral models of the TX IQ imbalance and PA distortions, as well as the behavioral models of other devices, channels and modules. In this way, the final output, and the outputs at the middle stages of the system in simulation,

can serve as a prediction of the performance of the signal in the real world applications, and show the influences of the TX IQ imbalance and PA at various stages of the communication system.

For the second task, digital signal predistortion is an efficient and low-cost method to compensate the distortion. After modeling the inverse function of the corresponding distortion, the digital predistorter is cascaded ahead of the distortion. The influence of the predistorter and that of the corresponding distortion are mutually neutralized, and the overall system is free of the specific distortion. The digital predistortion can eliminate the distortion at the transmitter side, avoid the possible interference on the adjacent wireless channels, improve the quality of the received signal, and simplify the signal processing complexity of the receiver. Digital predistortion can also increase PA efficiency and linearity without sacrificing output power level, hence reduce the energy consumption while maintaining high quality of the transmitted signal.

Behavioral model plays a critical role both in system-level simulation and in design of the digital predistortion. The quality of the behavioral model eventually determines the performance of both the system-level simulation and the digital predistortion. In this way, the research work of these two seemingly different subjects – system-level simulation and digital predistortion – are unified under one task: Behavioral Model.

Traditionally, the power amplifier and the TX IQ imbalance are treated, simulated and compensated differently and separately. It is partly due to their distinct and mutually independent physical characteristics. However, from the viewpoint of behavioral models, the inherent physical characteristics are ignored, while the corresponding device or system is viewed as a black box with some mathematical relationship between its input and output.

That motivates the research in this thesis on behavioral models treating TX IQ Imbalance and power amplifier jointly and as a single black box.

The advantages of treating these two distortions jointly and as a single black box are: (1) Theoretically, it gives a unified view of these two distortions, and makes it more abstract to facilitate system-level design and implementation. (2) Practically, it has the potential to save more resources than treating them differently and independently – less running time, less memory space, etc. (3) The predistortion methodology using a behavioral model at the transmitter can improve the transmitted signal quality and lessen the signal processing burden on the receiver, which still need to compensate other circuit and channel impairments. Hence it simplifies the design of the receiver and reduces the interferences to other wireless signals and systems at the same time.

Objective:

Specifically, the research includes two stages:

In the first stage, the behaviors of TX IQ imbalance and PA distortions are studied independently. The objective in this stage is mainly to improve the performances of the behavioral models of these two distortions individually, and make each meet the demand of the modern high-energy-efficiency and high-data-rate wireless communication systems. Specifically, the PA models should characterize the deep nonlinear and significant memory distortions, and the TX IQ imbalance models should characterize the significant memory effects. The behavioral models are used in digital predistortion to compensate the impairments. This stage is a stepping-stone to unify these two distortions in the next stage.

In the second stage, the objective is to unify the TX IQ imbalance and power amplifier distortion models into one black box, and characterize them in one unified behavioral

model. The unified model should have high accuracy, and strong capability to predict the performance of the signal impaired by these distortions. It should satisfy the requirements of the evolving modern wireless communication systems – deep nonlinearity and significant memory effects. It should also have simple implementation, and save resources. The unified model can be used in system-level simulation to predict the signal suffering these distortions; it can also be used to design the predistortion to improve the transmitter performances.

1.3 Major Contribution

1) An accurate and flexible test bed based on laboratory instruments is developed for characterizing and testing the PA behavioral models. This test bed can also be used to characterize and test TX IQ imbalance, and joint TX IQ imbalance and PA distortions with respective software modules. System impairments, such as TX IQ imbalance in the case of PA modeling, channel delay, frequency offset, and carrier phase offset, are all well compensated by digital signal processing methods in the test bed. The accurately recovered envelopes of the DUT's input and output signals are used for PA or TX IQ imbalance behavioral characterization and modeling. The competency of the test bed to accurately recover the envelopes of a PA's input and output signal is demonstrated through experimental results. The test system's EVM is lower than -41dB and relative envelope error (REE) is less than 5.5%. The results are published in the Author's Publication [3].

2) In the early stage of the author's work, a behavioral model based on memory polynomial model (MPM) is studied for characterizing and predistorting the power amplifier. It models (or predistorts) AM/AM and AM/PM separately by two different

memory polynomial models. On one hand, each memory polynomial model has limited number of items and coefficients (only the diagonal items of the Volterra model), and hence greatly reduces the influence of the condition number problem. While on the other hand, the two models are combined together to realize more items than a traditional memory polynomial model. Hence the combined model is more powerful and has more freedom to characterize or predistort the power amplifier distortion. Besides, by using AM/AM and AM/PM structure, the model also reflects the inherent physical characteristics of power amplifiers. The results are published in Author's Publication [1].

3) A new simple frequency-domain predistorter is proposed to compensate TX IQ imbalance of significant memory effects. The predistortion in frequency domain has the advantage of perfectly compensating significant memory effects in IQ imbalance, without introducing inter-symbol interference (ISI) for OFDM signal. By transforming the problem to a matrix problem, the predistorter is solved by robust numerical methods. Moreover, the implementation of the proposed predistortion scheme is simple and straightforward for OFDM systems.

4) The behavior of power amplifier distortion and TX IQ imbalance distortion are unified as a single black box, and characterized by a single model. A novel neural network – rectangular structured Focused Time-Delay Neural Network (RSFTDNN) – is proposed. Unlike the polynomial based model, which can only model mild nonlinear systems, the proposed RSFTDNN shows a strong ability and high accuracy in the unified model of the TX IQ imbalance and PA distortions. Two cases of experiments are carried on, where the RSFTDNN model shows better performance than the memory polynomial dual-input model. The results are published in Author's Publication [2].

1.4 Arrangement of the Thesis

In Chapter 2, the background information on the analog distortions of TX IQ Imbalance and power amplifier is firstly introduced. Then system-level simulation and digital predistortion of these two impairments are presented, where the quality of the behavioral model determines the performance. A comprehensive literature review on the behavioral models of these two impairments is carried out, e.g. Volterra model, memory polynomial model, and neural network, etc.

In Chapter 3, a general test bed is proposed to characterize the behavior of TX IQ Imbalance and power amplifier. It is based on laboratory instruments and uses several state-of-the-art methods to compensate the impairments in the channel. Firstly, the PA test bed structure is presented, followed by detailed steps and explanations of digital signal processing methods to compensate the various channel impairments. Then in the next two sections, the PA test bed is modified to facilitate the behavior characterization of TX IQ imbalance distortion, and that of joint TX IQ Imbalance and PA distortions, respectively. The performance metrics are also given, and the performance of the test bed is assessed by experiments.

In Chapter 4, a new behavioral model to characterize and predistort the power amplifier is proposed. It is based on the memory polynomial model, and has the structure to model or predistort AM/AM and AM/PM separately. It firstly introduces the memory influence on the AM/AM and AM/PM functions. Then the proposed new predistortion architecture using memory polynomial model is described, and its mathematical analysis and its predistortion block diagram are also presented. Simulation results are given, and compared with the traditional architecture.

In Chapter 5, a new simple frequency-domain predistorter is proposed to predistort TX IQ imbalance. Unlike the time-domain compensation methods, the proposed frequency-domain predistortion method can perfectly compensate long memory effects, without introducing any ISI impairment. Firstly, the IQ imbalance and *mirror frequency interference* is analyzed, and the frequency-domain IQ imbalance model is deducted. Then the drawback of time-domain IQ imbalance predistortion is analyzed. After that, the proposed frequency-domain IQ imbalance predistortion scheme is shown. The numerical issue is also considered, the noise influence and the problem of Normal Equation are analyzed, and some robust methods are proposed to identify the predistorter. Both simulation and real experiment results are shown.

In Chapter 6, the power amplifier distortion and TX IQ imbalance distortion are unified and characterized by a single neural network model. Firstly, the idea of unifying these two distortions is presented. Then the limitation of condition number on polynomial based models is analyzed. After that, the proposed neural network model for modeling PA-plus-TX-IQ-imbalance-distortion is shown. The overall experiment structure is illustrated, and the experiment results are given.

Finally, the conclusion and future work are presented in Chapter 7.

The Appendices provide information of the numerical issues discussed in the thesis.

Chapter 2 Literature Review

2.1 TX IQ Imbalance and PA Distortions

2.1.1 Power Amplifier Nonlinear Memory Distortion

Firstly, PAPR is defined as the ratio between the peak power level and the averaged power level of a signal [11]:

$$\begin{aligned} PAPR(x(t)) &= \frac{P_{peak}}{P_{average}} \\ &= \frac{\max_{0 \leq t \leq T} \{|x(t)|^2\}}{\frac{1}{T} \int_0^T |x(t)|^2 dt} \end{aligned} \quad (2.1)$$

where T is the time duration of the signal $x(t)$.

Power Amplifier is an energy generation device that increases the power level of the signal at the front end of the transmitter by several orders, so as to greatly increase the value of SNR (signal to noise ratio) of the receiver. However, PA is also the most energy-consuming device in the transmitter. Moreover, the power efficiency is low for systems with high PAPR signals. For example a signal with 10dB PAPR may require ideal Class A or Class B PAs having average efficiencies of 5% and 28%, respectively [12]. Working in nonlinear mode can significantly increase the efficiency. But it brings another problem – nonlinear impairments. Usually, designing a power amplifier demands a

tradeoff among the output power level, the efficiency and the degree of nonlinearity. And since energy-saving issue is more important, the linearity is often sacrificed, and the signal is suffering from nonlinear distortion, esp. for high PAPR signals like OFDM signals.

PA memory-effects-distortion also becomes significant, with the advent of wideband signals being widely deployed in the third (3G) and fourth (4G) generations of wireless communication systems. Thermal effect is one major cause of memory effects: the current temperature, which influences the PA output, is caused by both the current and the past energy inputs. Another important cause is the electrical charging and discharging, which is generated from both the designed and the parasitic capacitance. In the frequency domain, memory effects show up as frequency dependent distortion over the signal bandwidth.

The nonlinear and memory distortions of PA can be illustrated by its AM/AM and AM/PM relationships, as shown in Fig. 2-1 and Fig. 2-2. AM/AM depicts the relationship between the input signal power level and the output signal power level. Both power levels are denoted in decibels above 1mW, or dBm, as defined by:

$$P_{dBm} = 10 \cdot \log_{10}(P_{watt}) + 30 \quad (2.2)$$

AM/PM depicts the relationship between the input signal power level in dBm and the phase shift between output and input. Firstly, from AM/AM relationship, it is clear that as the input power level increases, the output power level increases linearly at low power level and compresses under high input power level as the PA is working under nonlinear mode. The degree of nonlinearity increases with power level for a given device. Please note that for a high PAPR signal, even although its average power level is in the linear zone, its peak power, which can be more than 10 times of the average level, can reach deeply into the nonlinear zone.

Secondly, the relationships of both AM/AM and AM/PM are not simple functions of input power level, as illustrated by the scattered points across a range of output power levels for a given input power level in Fig. 2-1 and Fig. 2-2 (data is acquired from the power amplifier ZVE-8G HPA). This is caused by the memory effects of the PA, i.e., the current output power level or phase shift is not only caused by current input signal sample, but the past input samples as well.

Fig. 2-1 AM/AM relation of power amplifier

Fig. 2-2 AM/PM relation of power amplifier

Due to the nonlinear distortion of the power amplifier, the OFDM signal loses its orthogonality among its subcarriers, and results in high BER. According to [13], after nonlinear distortion, the BER of OFDM signals does not decrease with the increase of SNR. Instead, there is a BER performance floor.

2.1.2 Transmitter IQ Imbalance Distortion

The overall diagram of the typical wireless transmitter is shown in Fig. 2-3. IQ imbalance distortion refers to the imbalance between the I-path and Q-path. The difference between the gains of the two paths is called *gain imbalance*. Another imbalance is called *phase imbalance*, i.e., the real phase difference between the I-path signal $x_{I,I}(t)$ and the Q-path signal $x_{I,Q}(t)$ are some degree away from 90° .

Just as the power amplifier, there are significant memory effects in the IQ imbalance distortion with wideband signals. It is partly owing to the reconstruction filters. And it shows frequency dependent IQ imbalance within the signal bandwidth.

Ideally speaking, the low-pass equivalent signal $x_l(t)$ of the RF signal $x_p(t)$ after the up-converter should be in the following relationship with the original transmitted signal $s(n)$ as:

$$x_l(n) = h_{TX}(n) \otimes s(n) \quad (2.3)$$

where $x_l(n)$ is the discrete form of $x_l(t)$, $h_{TX}(n)$ is the low-pass equivalent linear filter of the overall path from DAC to up-converter.

With TX IQ imbalance, the distorted signal $x_l(t)$ after up-converter is instead:

$$x_l(n) = h_{TX,d}(n) \otimes s(n) + h_{TX,m}(n) \otimes s^*(n) \quad (2.4)$$

where $h_{TX,d}(n)$ and $h_{TX,m}(n)$ are the complex-valued low-pass equivalent direct filter and mirror filter, respectively. Now there is an interference on the signal – the conjugate of the transmitted signal $s^*(n)$. It is called *Mirror Frequency Interference* [9].

Fig. 2-3 Overall diagram of a wireless transmitter, where the signal with the subscript p is the RF passband signal with carrier frequency, while the signal with the subscript l is the corresponding low-pass equivalent signal. The index n means the digital signal, while index t means the continuous analog signal after DAC. And θ_I and θ_Q are the phase imbalance owing to LO imbalance.

2.2 Behavioral Models of Transmitter Distortions

2.2.1 Behavioral Models in System-Level Simulation

Simulation is usually considered the indispensable step; and only after satisfactory simulation results having been obtained can a design be forwarded to implementation and test. Moreover, it is well known that simulation can speed up the overall design and implementation progress by simplifying the way to design and invent, analyzing the possible performance of a specific design, providing the intermediate results at various parts of a system, exposing potential problems or drawbacks at the early stage of a design, and predicting performance of the future products.

Generally speaking, the simulation in the field of wireless communication can be categorized into two classes: circuit-level simulation and system-level simulation. Circuit-level simulation caters for the design of a specific device. It is based on physical models or circuit-level models. These models mathematically describe the physical characteristics, effects, relationships and impairments. The advantage of circuit-level simulation and modeling are that they can show many details in the view of low-level design and manufacture, and hence very suitable for the design of a specific device or circuit. However, it requires heavy computation resources and a very long running time to accomplish the simulation of a single device. As a result, it cannot be used in the whole system simulation which contains many device models, some of which have significant memory effects.

For instance, the RF domain simulation, if in the circuit-level, will demand to compute a huge amount of data over even very short duration of signal. Since the RF domain

signals usually have carrier frequency of more than 1 GHz, and by virtue of Nyquist Theory [14], the circuit-level simulation needs a sampling rate of several giga samples per second – it needs to compute several giga samples of the signal of 1 second duration.

To circumvent the computation problem, system-level simulation is applied instead. From the viewpoint of system design, only the input and output of each device is concerned, while its physical characteristics and responses within the device are neglected. Hence in system-level simulation, each device is usually considered as a black box, with some mathematical relationship between its input and output.

Now for the RF signal, system-level simulation only concerns and simulates the envelope of the RF signal, while views its carrier frequency as an environment parameter and simply ignores it. The envelope of the RF signal is also called low-pass equivalent signal [14], which has only tens of megahertz. As a result, the sampling rate is reduced by several orders.

In a system-level simulation, to simplify the simulation and save computation resources, and by using low-pass equivalent signals in RF domain, different modules or devices are sometimes merged into one black box. For instance, the DAC and reconstruction filters, the mixer and its filter, the power amplifier and the TX antenna, are often treated as a single black box, respectively.

2.2.2 Behavioral Models in Digital Predistortion

Some of the TX analog distortions can be predistorted at the transmitter. It is done at the baseband digital domain by some processing methods, and often called *digital predistortion* (DPD), or *predistortion* for short. The prototype of the digital predistortion is

illustrated in Fig. 2-4. For a specific analog distortion, the predistorter is a behavioral model characterizing the inverse function of the given analog distortion. The transmitted digital signal is firstly predistorted in the digital domain, and then converted to analog signal by DAC, and passed through the analog devices. As a result, the inverse function of the distortion (predistortion) is cascaded with the distortion itself, and the overall effect is to remove the distortion. Since the digital predistorter is also a behavioral model, the numerous research results of the behavioral model in the literature can also be applied in predistortion.

The benefits of digital predistortion are: Firstly, it compensates the corresponding analog distortion from the signal, enhances the signal quality and reduces the complexity of signal processing at the receiver side. Moreover, it can prevent the spectral regrowth and impairments to the adjacent channels, and hence avoid the possible infringement of the spectrum requirements by the government. Finally, since it is done in the baseband digital domain, it is low cost and easy to implement especially in the terminal equipment. Hence digital predistortion is a highly efficient way to remove some of the analog distortions at the transmitter.

Fig. 2-4 Prototype of digital predistortion for analog impairments, where the discrete low-pass equivalent signals are used to represent the analog RF signals

2.3 Behavioral Models of Power Amplifier

The behavioral model characterizing power amplifier nonlinear and memory distortion has attracted intensive research interest over decades. Only the low-pass equivalent signals are considered. Henceforth, without further notice, the input and the output signals of a power amplifier always refer to the low-pass equivalent signals.

Traditionally, behavioral models of power amplifier do not take the memory effects into consideration. In the early stages of wireless communication, signals are usually narrowband. And power amplifiers approximately show memoryless nonlinear characteristics in the case of narrowband signals. Various memoryless behavioral models are proposed and applied to characterize the memoryless behavior of the power amplifier, e.g., Saleh models [15], Bessel-Fourier models [16, 17], Hettrakul and Taylor model [18] and Berman and Mahle model [19].

However, power amplifier nonlinear distortion is showing significant memory effects with the modern wideband signals in 3G and 4G wireless communication systems. Partly owing to the thermal effects, and partly owing to the charge and discharge of the circuits, the current output of the power amplifier is determined by both the current and past inputs. Hence, the current research is mainly focusing on characterizing both the nonlinearity and memory effects of the power amplifier.

PA behavioral models can be classified into polynomial based models and neural network models.

2.3.1 Polynomial Based Model

The major advantage of a polynomial based model is that it is in linear relationship with its coefficients. As a result, the identification of the polynomial based model can be done by using various linear numerical methods, e.g., least squares (LS) method, least mean squares (LMS) method and recursive least squares (RLS) method.

Volterra model [20] is widely seen as both the fundamental model and the typical representative. It is proposed by the Italian mathematician Vito Volterra in the year 1887 in his book *The Theory of Functionals* [21]. Volterra model can be viewed as the memory expansion of the Taylor series [22, 23], and “is particularly interesting as it produces the optimal approximation (in a uniform error sense) near the point where it was expanded” . And the discrete truncated version of Volterra model has been widely used to model various systems [20, 24, 25].

Ever since Norbert Wiener applied Volterra series to analyze nonlinear circuits in 1940s [26], various extensions and modifications of Volterra model have been proposed to simplify the complexity of the Volterra model. They are divided into two categories: nonlinear models with linear memory, and nonlinear models with nonlinear memory [27].

2.3.1.1 Nonlinear models with linear memory

The first category is to combine several sub-models to represent a general behavioral model. Wiener model and Hammerstein model [25, 28] are two typical models. Both use a nonlinear memory-less model to characterize the nonlinearity, and a linear time-invariant system to characterize the memory effects. The two sub-models are cascaded in serial to characterize PA's overall behavior – Wiener model (Linear-Nonlinear), and Hammerstein model (Nonlinear-Linear).

By increasing the number of sub-models, and varying the permutation of their orders, many types of structures can be implemented [27, 29-31]. One simple extension is the three-block model that adds one more linear memory block at the end of Wiener model – It is called Wiener-Hammerstein model [28]. The Poza-Sarkozy-Berger model [28] employs two Wiener-Hammerstein models to characterize the AM/AM and AM/PM relationships, respectively [32, 33]. The frequency-dependent Saleh model [15] characterizes the power amplifier distortions on the in-phase component and quadrature component of the signal by two Wiener-Hammerstein models. However, increasing the number of sub-models also increases the complexity of identification and simulation.

For this category of behavioral models, the implementation is simple, and the number of coefficients can be significantly reduced compared with the Volterra model. However, these models can only characterize linear memory effects. Another disadvantage is that since the nonlinear block is usually identified by single-tone signals at the center frequency, the model cannot predict the interaction between several tones [31]. Furthermore, as these models are nonlinear with respect to its coefficients, none of the linear identification methods can be used. And identifying these models can become very complicated and unreliable [34].

2.3.1.2 Nonlinear models with nonlinear memory

In the second category, the behavioral models can characterize the nonlinear memory effects. The nonlinear memory effects are mainly caused by the active-device low-frequency dispersion, electro-thermal interactions, and bias circuitry [27, 35-40]. These models are usually modified and/or simplified Volterra models.

The straightforward way to simplify the Volterra model is to directly prune its items. Among the various pruning methods, memory polynomial model [41] is one of the simplest model. It removes all the other items, and only retains the “diagonal” items of the Volterra model. Later in [34], a more general memory polynomial model is proposed by adding some “near-diagonal” items.

Dynamic pruning is to divide the Volterra model into two parts: static and dynamic parts. After that, it prunes the two sets of items in different ways, according to their different characteristics [31, 42, 43]. Dynamic pruning model can separate the purely static effects from the dynamic effects, which are inherently blended in the classical Volterra series [31].

Another effort to modify the Volterra model is Spline-based model [44, 45]. Spline function is in the form of piecewise polynomials, where the amplitude of the input signal is divided into several intervals. Since the amplitude is required to be smooth and continuous, the function and its derivatives should be continuous at the joint points.

The models of non-orthogonal basis suffer the problem of convergence [46]. There is a strict limitation on the range, in which the non-orthogonal basis model can converge. The limitation on the range of convergence can be largely relaxed by using orthogonal bases. Several time-domain orthogonal models are proposed [47-51]. Kautz-Volterra (KV) model [48, 49] is a typical orthogonal model that is based on nonlinear IIR structure and feedback loops. Laguerre function [50, 51] is another orthogonal basis used to model power amplifiers.

All the above mentioned models are in time-domain. There are also frequency-domain PA behavioral models [52-54]. Since each frequency component of the input signal is

orthogonal to one another, the frequency-domain behavioral models are of orthogonal basis. And the implementation is especially convenient to OFDM communication system – Since the OFDM system already contains IFFT operation in the transmitter and FFT operation in the receiver, there is no need for any extra transformation.

There is a question on the convergence [46] of the Volterra series based models. And by pruning a significant number of items of the original Volterra model, how much pruning-error is generated? Up to now, there is no theoretical answer to these questions. Moreover, even though the condition numbers of the simplified models are considerably reduced, these condition numbers still increase quickly with the increase in nonlinear order and memory depth, resulting degradation of the accuracy of model identification.

In the following section, we are going to review the Volterra model, which is the fundamental model of polynomial based models. Then the memory polynomial model is reviewed, which is widely used and one of the simplest models modified from the Volterra model. Finally, the counterpart of polynomial based models – the neural network is also reviewed.

2.3.2 Volterra Model

Volterra model is a Taylor expansion with memory items, and suitable to model a weakly nonlinear time-invariant system with fading memory [23]. Volterra model is linear to its coefficients and thus can take advantage of linear identification algorithms.

In continuous time form, the Volterra model with K denoting the highest nonlinear order is shown [46] as:

$$y(t) = \sum_{k=1}^K y_k(t) \quad (2.5)$$

and each specific-nonlinear-order-output is:

$$y_k(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\tau_1, \dots, \tau_k) s(x - \tau_1) \cdots s(x - \tau_k) d\tau_1 \cdots d\tau_k \quad (2.6)$$

where $s(t)$ is the input signal, $y(t)$ is the output signal.

The corresponding discrete time form, with K denoting the highest nonlinear order and M denoting the memory depth, can be written as:

$$y(n) = \sum_{k=1}^K y_k(n) \quad (2.7)$$

And each specific-nonlinear-order-output is:

$$y_k(n) = \sum_{m_1=0}^{M-1} \cdots \sum_{m_k=0}^{M-1} h_k(m_1, \dots, m_k) \prod_{i=1}^k s(n - m_i) \quad (2.8)$$

If the Volterra coefficient $h_k(m_1, \dots, m_k)$ is same for any permutation of the time delay set $(m_1, \dots, m_k)$, the coefficient $h_k(m_1, \dots, m_k)$ is called symmetric, and the number of items in (2.8) can be reduced significantly by combining the items of the same coefficients:

$$y_k(n) = \sum_{m_1=0}^{M-1} \sum_{m_2=m_1}^{M-1} \cdots \sum_{m_k=m_{k-1}}^{M-1} h_k(m_1, \dots, m_k) \prod_{i=1}^k s(n - m_i) \quad (2.9)$$

Both the above continuous time form and discrete time form of Volterra model are only applicable to passband RF signals. To characterize the power amplifier behavior, the corresponding low-pass equivalent Volterra model should be developed. Consider the relationship between the passband signal $y(t)$ and its low-pass equivalent signal $y_l(t)$:

$$\begin{aligned} y(t) &= \Re\{y_l(t)e^{j2\pi f_0 t}\} \\ &= \frac{1}{2}\{y_l(t)e^{j2\pi f_0 t} + y_l^*(t)e^{-j2\pi f_0 t}\} \end{aligned} \quad (2.10)$$

Replacing (2.10) into (2.6), and filtering out all the harmonic frequencies, the continuous time low-pass equivalent Volterra model turns out to be [28]:

$$y(t) = \sum_{k=1}^K y_k(t) \quad (2.11)$$

where

$$y_k(t) = \begin{cases} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h_k(\tau_1 \cdots \tau_k) \prod_{r=1}^{(k-1)/2} s^*(x - \tau_r) \prod_{r=(k+1)/2}^k s(x - \tau_r) d\tau_1 \cdots d\tau_m & k \text{ is odd} \\ 0 & k \text{ is even} \end{cases} \quad (2.12)$$

Note that all of the even nonlinear order items are removed.

And the corresponding discrete time form of the low-pass equivalent Volterra model is:

$$y(n) = \sum_{\substack{k=1 \\ k \text{ is odd}}}^K y_k(n) \quad (2.13)$$

where the Volterra kernel [46] is:

$$y_k(n) = \sum_{m_1=0}^{M-1} \sum_{m_2=m_1}^{M-1} \cdots \sum_{m_k=m_{k-1}}^{M-1} h_k(m_1, \dots, m_k) \prod_{i=1}^{(k-1)/2} s^*(n - m_i) \prod_{r=(k+1)/2}^k s(n - m_r) \quad (2.14)$$

Volterra model can only characterize mild nonlinear system [25, 27]. The reasons are: Firstly, the series does not converge for strong nonlinearity [28]. Secondly, the complexity – the number of the coefficients – increases dramatically with the highest nonlinear order and memory depth [55]. And thirdly, it is very difficult to accurately identify the coefficients of a high nonlinear order or a long memory depth. The last issue is owing to the fact that the condition number is increasing very fast with the highest nonlinear order or memory depth.

2.3.3 Memory Polynomial Model

One way to simplify Volterra model is to directly prune the items, and many research results have been published to reduce the number of Volterra model items in one way or another [31, 56, 57]. In [41], a simplified Volterra Model – memory polynomial model – is proposed. The model is in the following form:

$$y_n = \sum_{m=0}^{m=M} B_m(b_m, s_{n-m}) \quad (2.15)$$

where M is the memory depth in unit of samples, and

$$\begin{aligned} B_m(b_m, s_{n-m}) &= s_{n-m} \cdot \sum_{k=1}^K b_{mk} |s_{n-m}|^{k-1} \\ &= s_{n-m} \cdot \beta_m(|s_{n-m}|) \end{aligned} \quad (2.16)$$

where K is the order of polynomial $B_m(b_m, s_{n-m})$.

Later in [58], the memory polynomial model is expressed in an alternative and clearer way:

$$y(n) = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} h_{km} s(n-m) |s(n-m)|^k \quad (2.17)$$

In [31], the authors analyze the memory polynomial model, and call it “probably the simplest non-trivial pruned Volterra model”. The memory polynomial model only uses the “diagonal” items of the Volterra model, i.e. $m_1 = m_2 = \dots = m_k$ in (2.14). The number of coefficients is greatly reduced. Meanwhile, this model can be easily implemented, with the structure similar to a finite impulse response (FIR) filter, only replacing the linear gain by higher nonlinear order polynomials [31].

However, it is often too simple to be “powerful” enough to accurately model a power amplifier. Later, researchers tried to extend the model to have a better performance.

In [56], the unit delay tap in the memory polynomial model is changed into sparse delay tap. By this way, it provides more freedom to characterize the memory effects by using variant-length time space between every two adjacent taps. And a longer memory effects can be modeled with still relatively small amount of coefficients. The form of the memory polynomial model with sparse delay taps is given by:

$$y[n] = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} h_{km} [s(n - d_m^{(M)})] |s(n - d_m^{(M)})|^k \quad (2.18)$$

where $d_m^{(M)}$ is the sparse delay tap and the number of taps is M [56].

The “diagonal” restriction reduces tremendously the number of the coefficients [31]. However, the usage of only “diagonal” items also limits the accuracy, since some of the “off-diagonal” items may sometimes be more important than the “diagonal” ones [31].

In [57], a “near diagonal” structure is proposed. The “near diagonal” items mean $|m_i - m_j| \leq l$ in (2.8), where l is a small integer [57]. By increasing the value of l , more “near diagonal” items are included, and the normalized mean square error (NMSE) can be improved [57]. The disadvantage is that since more items are included in the model, the complexity to extract the coefficients is also increased. In general, the “near diagonal” structure gives a way to tradeoff between the model accuracy and its complexity by varying the value of l [57].

The “near diagonal” structure can be expressed as three parts: traditional memory polynomial, its positive cross-term time shifts, and its negative cross-term time shifts. Accordingly, a general memory polynomial model with cross-terms can be written as [34]:

$$\begin{aligned}
y_{GMP}(n) = & \sum_{k \in K_a} \sum_{l \in L_a} a_{kl} s(n-l) |s(n-l)|^k \\
& + \sum_{k \in K_b} \sum_{l \in L_b} \sum_{m \in M_b} b_{klm} s(n-l) |s(n-l-m)|^k \\
& + \sum_{k \in K_c} \sum_{l \in L_c} \sum_{m \in M_c} c_{klm} s(n-l) |s(n-l+m)|^k
\end{aligned} \tag{2.19}$$

where $K_a, L_a, K_b, L_b, M_b, K_c, L_c, M_c$ are the specifically chosen orders for the general memory polynomial model.

In [53], a time domain behavioral model is used to predistort power amplifier, whose coefficients are estimated in frequency domain. The advantage is that it is easier to compensate the delay effects of the transmission path, and align the transmitted and received signals.

2.3.4 Neural Network Based Model

Besides polynomial based models, there is another category of power amplifier behavioral model – neural network [59]. It has several inner stages and many inner-states, and the final output is determined by both the input and the inner states. Here a state denotes the output of a neuron, while a stage denotes a layer containing many neurons.

Fig. 2-5 A typical Time-Delay Neural Network (TDNN) with 2 inner layers

Time delay neural network (TDNN) is proposed to model PA behavior [59-63]. It is based on feed-forward neural network structure [60]. The input is delayed at the first layer, and then feed forwardly into the network. A typical structure is shown in Fig. 2-5. The input signal is delayed at the input layer, where z^{-1} means one sample delay. $f(\cdot)$ is the activation function. And $x_{i,(l)}(n)$ is the inner state of Neuron i and Layer l (inner stage l). And N_l denotes the total number of neurons of Layer l .

In [62, 63], a TDNN neural network is proposed to predistort the power amplifier, and the performance of convergence rate is compared among various adaptive identification methods. The neural network is composed of 2 inner layers, with the activation function being tansig function.

Recurrent neural network (RNN) [64-66] is a feedback-based neural network structure. The advantage of RNN structure is that it has the potential to model very long memory

effects with much less layers and neurons. However, due to the feedback branch, RNN has the risk of instability. Fig. 2-6 shows a typical RNN structure.

Fig. 2-6 A typical Recurrent neural network (RNN) with 2 inner layers and single output.

The input signal is delayed at the input layer, where z^{-1} means one sample delay. $f(\cdot)$ is the activation function.

In [67], a radical-basis function neural network (RBFNN) is proposed.

Fig. 2-7 is the overall structure of RBFNN. The input nodes $\vec{x} = [r_n, r_{n-1}, \dots, r_{n-L}]$ are the current and past input signals, therefore introduce the memory into the model. And φ_i is the Green function:

$$\varphi_i = G(\|x - t_i\|) = \exp\left(-\frac{\|x - t_i\|^2}{2\sigma_i^2}\right) \quad (2.20)$$

Fig. 2-7 RBFNN with $L + 1$ input nodes, M hidden nodes, and two output nodes. The two output nodes correspond to the AM/AM and AM/PM distortion, respectively [67].

Spatiotemporal neural network is proposed in [68] to predistort power amplifier nonlinearity. The neural network interpolates adjacent neurons, and prevents discontinuities among neurons of different layers. The drawback is that spatiotemporal neural network requires more neurons than TDNN to achieve certain level of normalized mean square error (NMSE) [68]. And a neural network without delay is also proposed [69] to model the circuit-level characteristics of the power amplifier, and is reported to achieve a high accuracy.

Neural network has the strong capacity to model deeply nonlinear and memory systems [27, 70, 71]. With a proper number of layers and neurons, the neural network can

characterize a system with arbitrary accuracy [70]. Neural network has variant number of inputs and outputs by its nature, and can model double input system quite straightforwardly. Last but not least, neural network enjoys more freedom, i.e., it can vary the number of layers and neurons in each layer, or change the type of activation function, or even change the connections or the whole structure.

The drawback of neural network is that its coefficients are not in a linear relationship with its output. Hence it cannot be identified by linear methods. A well-known identification method called *back-propagation* (BP) is widely used to extract its coefficients [70].

2.4 TX IQ Imbalance and Behavioral Models

Fig. 2-8 Classification of IQ imbalance compensation methods: (a) pre-compensation for TX IQ imbalance. (b) post-compensation for RX IQ imbalance. (c) post-compensation for joint TX and RX IQ imbalance.

Both the transmitter and the receiver suffer IQ imbalance distortions. OFDM signals are very sensitive to IQ imbalance distortion. Numerous compensation methods have been published. According to their location and scope, these methods can be classified into three groups: pre-compensation at the transmitter to correct TX IQ imbalance, post-compensation at the receiver to correct RX IQ imbalance, and post-compensation at the receiver to correct joint TX and RX IQ imbalance. They are illustrated in Fig. 2-8. Post-compensation for RX IQ imbalance, which has attracted most of research interest, is firstly reviewed. Then as a natural extension, post-compensation for joint TX and RX IQ imbalance is reviewed. Finally, pre-compensation for TX IQ imbalance, which is the subject of this thesis, is reviewed as the last category. Note that *pre-compensation* is also called *predistortion* in this thesis.

2.4.1 Post-Compensation for RX IQ Imbalance

The methods to compensate RX IQ imbalance at the receiver are based on the assumption of ideal transmitter, i.e., there is neither IQ imbalance nor any other impairments at the transmitter. These methods can be further categorized into three groups: calibration-based methods, reference symbol methods, and blind compensation methods.

Calibration methods [72] operate in off-line mode, and use some known calibration signal to estimate the value of RX IQ imbalance. After that, the calibrated value is used to process and correct the received signal to eliminate RX IQ imbalance distortion. In the early stage, frequency-independent or memory-less RX IQ imbalance is calibrated and compensated [73-76]. With the advent of OFDM and other wideband signals, methods are published to calibrate RX IQ imbalance of significant memory effects (or so-called frequency-dependent imbalance) [77, 78].

Reference symbol methods can be seen as on-line calibration methods, which estimate and compensate RX IQ imbalance by transmitting some known symbols. One common reference symbol is the pilots [79-84]. Recently, some pilot-aided methods are proposed to jointly estimate RX IQ imbalance and another physical impairment, e.g. carrier frequency offset (CFO) [80-82]. Another commonly used reference symbol is the training sequence [85-88], e.g. the short and long preambles in IEEE802.11a WLAN standard.

In order to avoid using known calibration symbols or wasting the spectrum resource, blind compensation methods are proposed. It mainly explores some statistical properties of the received signal, and uses the statistical information to adjust or compensate the signal that suffers RX IQ imbalance distortion. These methods can be classified as *blind adaption* [89-91] and *blind estimation* [90, 92, 93]. Blind adaption does not need to estimate any parameter, but to build an adaptive filter, and compensate the distorted signal by passing it through the filter. On the other hand, blind estimation is to estimate some parameters relating to RX IQ imbalance, and use them to compensate the signal.

For all the three groups of post-compensation methods, the assumption that the transmitter is perfect and without any impairment is too ideal. Hence by loosening the requirements and assuming that the transmitter also has IQ imbalance, more complex post-compensation methods are proposed to compensate joint TX and RX IQ imbalance.

2.4.2 Post-Compensation for joint TX and RX IQ Imbalance

The way to post-compensate joint TX and RX IQ imbalance is usually based on some reference symbol; hence it can be seen as a generalization of the reference symbol methods in the previous section. In [58, 94-98], joint TX and RX IQ imbalance distortions are compensated for SISO OFDM systems by the aid of reference symbols. And

compensation methods for MIMO OFDM systems are also proposed as an extension [7, 99-103].

Recently, blind or semi-blind methods [104-106] are published for TX/RX IQ imbalance compensation. A semi-blind method with the aid of a reference signal is proposed in [104]. In [105], the TX/RX IQ imbalance over a doubly selective (time-varying and frequency-selective) channel is compensated by a semi-blind compensation method with the aid of pilots. In [106], a blind method is to jointly compensate TX/RX IQ imbalance for MIMO OFDM over a doubly selective channel.

Several frequency-domain compensation methods are also proposed for IQ imbalance. A frequency-domain equalization method is proposed in [107] to jointly compensate IQ imbalance, CFO and IBI. To jointly compensate CFO and IQ imbalances, the time-domain compensator and frequency-domain compensator are iteratively and adaptively updated to achieve high accuracy in [108]. In [109], differential filter is working together with IQ imbalance estimator to compensate both DC and IQ imbalance impairments.

One of the problems of these post-compensation methods is that they have to neglect some other physical impairments and make some ideal assumptions to avoid making things too complicated. Besides TX/RX IQ imbalance, there are many other impairments, e.g. carrier frequency/phase offset, DC offset, time synchronization error, phase noise, time-varying and frequency-selective channel, PA nonlinearity. Although some efforts are devoted to jointly compensate TX/RX IQ imbalance and one of the other impairments [96, 105, 106], there is no available method that can consider all the impairments and still compensate IQ imbalance – it is too complex to include every impairment, and too heavy a burden for the receiver to compensate them all. Hence the practical solution is to

compensate locally – compensate RX IQ imbalance at the receiver, and compensate TX IQ imbalance at the transmitter. The former one has been discussed in the previous section, while the latter one is to be reviewed in the next section.

2.4.3 Predistortion of TX IQ Imbalance

Fig. 2-9 demonstrates the idea of TX IQ imbalance predistortion. The predistorter is the inverse model of the TX IQ imbalance. By cascading the predistorter with the TX IQ imbalance, the overall effect is that there is no TX IQ imbalance distortion in the output RF signal $y(t)$. Fig. 2-10 shows the implementation diagram. The predistorter is placed in digital domain (before DAC). Note that the predistorter of the TX IQ imbalance is a behavioral model. Hence the quality of the behavioral model and the accuracy of its identification eventually determine the performance of TX IQ imbalance predistortion.

Fig. 2-9 Prototype of digital predistortion of TX IQ imbalance, where $y_l(n)$ is the discrete low-pass equivalent signal of the RF signal $y(t)$ after up-converter

Fig. 2-10 TX IQ Imbalance Predistortion

Traditionally, TX IQ imbalance is treated as memoryless distortion (*frequency-flat* or *frequency-independent*). As a result, its predistorter is implemented by some memoryless models [110-114]. With the wideband signals in modern communication systems, TX IQ imbalance distortion shows significant memory effects (*frequency-dependent*), and is modeled (or predistorted) by some linear memory system [115-117]. In [117], the TX IQ imbalance are modeled by four parallel linear filters passing or crossing I and Q-paths, as illustrated in Fig. 2-11. All four frequency dependent branches are real-valued linear filters. $h_{II}(n), h_{QQ}(n)$ are the equivalent passing-through filters of I-path and Q-path, respectively. $h_{IQ}(n), h_{QI}(n)$ are the equivalent cross-talking filters between I-path and Q-path.

Fig. 2-11 Behavioral model of the TX IQ imbalance distortion [117]. And

$\alpha_{II}, \alpha_{IQ}, \alpha_{QI}, \alpha_{QQ}$ are the weights for each path.

The advantage of predistortion is that since it is implemented in digital baseband domain before the DAC, the TX IQ imbalance predistorter can be easily implemented either in microprocessor, or integrated circuits (IC). The current predistorters are

implemented by some time-domain behavioral models. However, there are some shortcomings in the time-domain behavioral models.

One problem is that the time-domain memory models may result in inter-symbol interference (ISI) even for OFDM signals. Usually each OFDM symbol has a cyclic-prefix as a guard interval between its adjacent symbols in order to avoid ISI. If the memory depth of the predistorter is longer than the length of the cyclic-prefix, the guard interval will be violated, resulting in ISI degradation. Hence the length of the cyclic-prefix set the limitation of the longest memory effects that the predistorter can compensate.

Even if the length of the predistorter is less than the length of cyclic-prefix, TX IQ imbalance predistorter can still cause ISI impairment. It has the effect of “shortening” the length of the cyclic-prefix – The TX IQ imbalance predistorter can “work” together with the other memory devices and modules at the later stages, such as the reconstruction filters, the multipath channel, and the filters at the receiver, etc. Consequently, the length of the equivalently overall memory effects (combining all the memory devices and channel throughout the whole transmission path) is very likely to grow longer than the length of the cyclic-prefix. And hence ISI is still generated, and BER is degraded subsequently.

Another consideration for the TX IQ imbalance predistorter is the *condition number* problem. Even for a linear memory model, more memory depth means more number of correlated items and thus higher condition number. Hence a TX IQ imbalance predistorter of many time-delayed items can suffer from mild *condition number* problem, which could compromise the accuracy of the predistorter.

Recently, pilots are proposed to estimate the predistorter in the frequency domain [118, 119]. The basic idea is to use the transmitted pilot and its conjugate to calculate the

coefficients of the predistorter. However, the drawback is that the mathematical equations are complex, and the noise and error can be accumulated and amplified throughout the whole process. Hence the methods are sensitive to noise, and lack of robustness.

2.5 Current Results on Unified Model of PA + IQ Imbalance

Recently, some polynomial based models have been proposed to treat PA distortion and TX IQ imbalance in a single unified model. In [120], the Hammerstein based model is proposed to predistort joint impairments of PA and IQ imbalances. In the scheme, the input signal and its complex conjugate are predistorted separately before combined together to form the final output. Since it contains a set of paralleled Hammerstein model, the structure of the predistorter is complex.

Instead predistorting the input signal and its conjugate separately, H. Cao et al. in [121, 122] propose to predistort the I-path, Q-path and their cross-talking paths of the input signal separately, and then combined them to form the final output. The scheme is call *dual-input* model because it contains two inputs – I-path input and Q-path input. Both Volterra model and memory polynomial model are proposed for the scheme, where memory polynomial based dual-input model has the advantage of much simpler form and fewer coefficients.

However, the abovementioned polynomial based models can only model mild nonlinear system [25, 27], because it is difficult to accurately identify the coefficients with high nonlinear order, owing to the *condition number* problem [123].

The neural network has also been applied to model the behavior of power amplifier nonlinearity and memory effects, either in feed-forward structure [59-63, 68] or in

feedback [64-66] structure. However, these models do not consider IQ imbalance. In [124, 125], a TDNN neural network is proposed to jointly predistort PA nonlinearity and IQ imbalance impairments. The neural network is a kind of real-valued TDNN model, and uses *hyperbolic tangent* function as its activation function. Its structure is fixed at 2 inner layer, which may limit the freedom and modeling capacity of the neural network. Meanwhile, spatiotemporal neural network is also proposed to linearize the power amplifier in the presence of IQ imbalance [126]. However, spatiotemporal neural network is reported to require more neurons than its TDNN counterpart.

2.6 Conclusion

Behavioral models on two key transmitter analog distortions – power amplifier nonlinear and memory distortion, and TX IQ imbalance distortion – can be used both in system level simulation for the sake of various algorithms design, and in predistortion to compensate these impairments at low cost. As shown in the literature review in this chapter, current behavioral models and compensation methods suffer from some shortcomings. It is usually analyzed on a case-by-case basis, and tradeoffs have to be made. Hence more efforts are needed to find behavioral models that are accurate, simple to implement, easy to identify, and without generating extra impairments.

Chapter 3 Test Bed for TX IQ Imbalance and Power Amplifier Behavioral Characterization and Modeling

3.1 Introduction

An accurate and flexible test bed is critical to TX IQ Imbalance and PA behavioral characterization and modeling. It should provide sufficient facilities to engage and complete the processes of characterization of the behavior of power amplifiers and TX IQ imbalance, identification of various behavioral models, and test and verification of the identified models. And it should be open and flexible to implement different signal recovery methods, and with no limitation on the type of input signals.

Since PA AM/AM and AM/PM characterization and behavioral modeling are only concerned with the envelopes of the corresponding RF signals, the test bed is required to accurately recover the envelopes of the PA's input and output signals in the form of digital samples. The test bed requires almost all the functions of a physical receiver, e.g., down conversion, ADC, demodulation, and baseband signal processing. Moreover, it should perfectly compensate all kinds of impairments occurring in the channels while leaving the PA nonlinear memory influence untouched, e.g., channel delay, frequency offset, or IQ imbalance, etc. Meanwhile, the test bed also needs a flexible signal generator which can

output enough linear power to drive the PA and generate all kinds of signal forms, such as two-tone, OFDM, and white Gaussian signals, to facilitate the PA behavioral characterization and modeling.

Similar considerations are also applicable for the test bed of TX IQ imbalance, and the test bed of joint TX IQ Imbalance and PA. There are only two differences from PA test bed: 1) The IQ imbalance in the signal generator is not considered as impairment anymore, but the very device under test (DUT) that needs to be modeled. 2) The input of the behavioral model of TX IQ imbalance is not an envelope of any RF signal, but the baseband transmit signal that can be conveniently and accurately acquired from the baseband part of the transmitter.

Currently a typical test bed is based on a specialized instrument called vector signal analyzer (VSA) [127-129], often equipped with a specialized digital signal analysis software, e.g., Agilent 89600 VSA software. The VSA instrument receives the RF signal, down-converts it to intermediate frequency (IF), executes ADC and digital demodulation. The acquired baseband signal is then passed to a PC in complex form, where signal analysis software is running to digitally process the signal. Meanwhile, real-time spectrum analyzer (RSA) works in the similar way as the VSA [130], and can serve as the receiver in a test bed. However, the major drawback is their high price. Moreover, all the functions are packed into the instrument and embedded analysis software. Little modification can be made, rendering it inflexible or unable to work for scenarios which require different test methods to compensate for the various kinds of impairments present in the test bed.

Vector Network Analyzer (VNA) is also used as a test bed to model PA behavior [131]. However, traditional VNA can only measure the static behavior due to its long sweep time,

hence it is unable to measure PA's instantaneous or dynamic behavior [127]. Large signal network analyzer (LSNA) is used in [132] to model the PA behavior over a wide range of frequency span. However, Fager et al. [133] argues that LSNA requires the modulation to be periodic, which limits the type of testing signals. Moreover, whether it can measure PA's instantaneous behavior is still unknown.

Instead of using specific instruments or software, the oscilloscope or other general purpose data acquisition systems, which works as analog-to-digital converter with a proper RF bandwidth and maximum sampling rate, is a good alternative for a TX IQ Imbalance and PA behavior test bed. If the RF bandwidth of the oscilloscope is higher than the signal's carrier frequency, the signal can be acquired directly; otherwise, an external down-converter is required. It turns out to be low cost, more flexible to implement different signal recovery methods, and with no limitation on the type of input signals.

There are several PA behavior test beds based on oscilloscope. In [134], the signal is directly down-converted to baseband before acquired into oscilloscope. In [133] the signal is directly acquired without external down-conversion, assuming the RF bandwidth of the oscilloscope is higher than the signal's carrier frequency. However, none of them give details on how the test bed is implemented, nor on compensation of different kinds of impairments along the transmission path. The authors in [135] show a detailed WiMAX system test bed based on oscilloscope and the steps to implement and compensate various impairments. However, the test bed is only for WiMAX system quality measurement, yet not for PA or TX IQ imbalance behavioral characterization and modeling.

In this chapter, detailed steps on how to implement the TX IQ Imbalance and PA behavior test bed based on general purpose oscilloscope is shown. The major impairments

such as IQ imbalance (for PA test bed), channel delay, frequency offset and phase offset, etc., are considered, analyzed and compensated. After compensating all the impairments, the envelopes of the DUT input and output signals are accurately recovered in the form of digital samples for behavioral modeling.

The structure of the chapter is organized as follows: In Section 3.2, the overall PA test bed structure is presented, followed by detailed steps and explanations of software signal processing implementation to compensate the channel impairments. Section 3.3 and Section 3.4 show how to set up TX IQ imbalance test bed and joint TX IQ Imbalance and PA test bed by modifying the PA test bed proposed in Section 3.2. In Section 3.5, quantitative performance metrics are given. In Section 3.6, the hardware design is shown to provide channels of both high SNR and high linearity. In Section 3.7, the performance of the test bed is assessed by experiments, and the results are presented. And Section 3.8 gives the conclusion.

3.2 Proposed PA Test Bed and Signal Processing

3.2.1 Test Bed Overall Structure

Fig. 3-1 shows the overall structure of the test bed. Here the RF bandwidth of the oscilloscope is assumed to be lower than the carrier frequency of the RF signals, so two external down converter channels are used to convert the signals to intermediate frequency. If the RF bandwidth of the oscilloscope is high enough, the external down conversion can be omitted. Throughout the chapter, it is assumed that external down converters are required.

Agilent E4438C ESG is a vector signal generator that can generate arbitrary waveform and up-convert it to RF signal. In cases where the RF frequency of the signal generator is lower than the required carrier frequency, up-converter may be used. The RF signal power level should be high enough to drive the device under test (DUT: high power amplifier) to work in its nonlinear zone. The input and output signals of the DUT pass through separate down conversion channels – Channel A and Channel C. In each channel, the signal passes through an attenuator before down conversion. This ensures that the input signal level presented to the down-converter is in its linear zone.

Fig. 3-1 Overall Structure of Test Bed for PA behavioral modeling

Finally, the PA input and output signals in IF frequencies, denoted as Sig A and Sig C, are acquired by two channels of the oscilloscope. The signals from the oscilloscope are digitized and are passed to a personal computer (PC) for digital demodulation,

impairments compensation, and other digital signal processing. The accurately recovered DUT input and output signals in baseband digital forms are used to characterize and model the DUT behavior.

3.2.2 Software Overall Structure

Fig. 3-2 Block diagram of digital signal processing

The overall structure for digital signal processing which is implemented in software running on the PC is shown in Fig. 3-2. At the transmitter, the baseband waveform of the test signal is generated, followed by IQ imbalance linear predistortion. At the receiver, channel impairments such as channel delay, frequency offset, and phase offset are compensated.

3.2.3 TX IQ Imbalance Compensation

In our test bed, IQ imbalance in the demodulator is avoided by the use of heterodyne structure and digital demodulation. But IQ imbalance in the modulator exists in the Agilent E4438C ESG VSG due to its analog direct up conversion structure. Although they do not influence the relationship between the envelopes of the PA's input and output signals, where they suffer from the same IQ imbalance, IQ imbalance in modulator do degrade the performance of PA behavioral characterization and modeling in two ways. Firstly, TX IQ imbalance may severely distort the training sequences, and in turn degrade the performance of time alignment, frequency offset compensation and phase offset compensation. As a result, the accuracy of the recovered envelopes is significantly decreased. Secondly, IQ imbalance may degrade the evaluation metrics such as adjacent channel power ratio (ACPR) and error vector magnitude (EVM). In this case, the metrics cannot precisely evaluate the performance of PA behavioral modeling and linearization, because part of the degradation is due to IQ imbalance in the modulator instead of the PA nonlinear and memory effects. Hence the frequency-dependent IQ imbalance in the modulator should be eliminated.

A dual-input linear predistorter is proposed by L. Ding et al. [117] to compensate IQ imbalance in a similar way that a nonlinear predistorter linearizes a nonlinear PA. The

linear predistorter only requires to inversely model the IQ imbalance using linear FIR filters. The method is shown in Fig. 3-3.

Fig. 3-3 Linear predistorter to compensate IQ imbalance

The IQ imbalance including crosstalk is modeled by 4 FIR filters, and the linear predistorter is also modeled by another 4 FIR filters correspondingly, as shown in Fig. 3-4.

Fig. 3-4 IQ imbalance model and linear predistorter

In Fig. 3-4, the complex signal $u(n)$ is the original signal, $x(n)$ is the predistorted signal that goes into DAC, and $y(n)$ is the low-pass equivalent of the RF signal after up conversion. These signals can be expressed as:

$$\begin{cases} u(n) = u_i(n) + j \cdot u_q(n) \\ x(n) = x_i(n) + j \cdot x_q(n) \\ y(n) = y_i(n) + j \cdot y_q(n) \end{cases} \quad (3.1)$$

The 4 FIR filters in the IQ linear predistorter may have different lengths of memory, i.e. the number of taps. To simplify the expression, here it is assumed that the 4 FIR filters have the same length M (For the case of non-equal-length filters, the lengths of filters can be made equal by padding zeros to the shorter ones, and M is the length of the longest filter). By using vector expression, a sequence of data $x(n)$ and $u(n)$ can be written as:

$$\mathbf{x} = \mathbf{U}_i \mathbf{p}_i + \mathbf{U}_q \mathbf{p}_q = [\mathbf{U}_i, \mathbf{U}_q] \begin{bmatrix} \mathbf{p}_i \\ \mathbf{p}_q \end{bmatrix} \quad (3.2)$$

where

$$\begin{cases} \mathbf{x} = [x(M-1), \dots, x(N-1)]^T \\ \mathbf{p}_i = [p_i(0), \dots, p_i(M-1)]^T \\ \mathbf{p}_q = [p_q(0), \dots, p_q(M-1)]^T \end{cases}$$

with

$$\begin{cases} p_i(n) = p_{11}(n) + j \cdot p_{21}(n), & 0 \leq n \leq M-1 \\ p_q(n) = p_{12}(n) + j \cdot p_{22}(n), & 0 \leq n \leq M-1 \end{cases}$$

and $\mathbf{U}_i = \mathbf{Re}\{\mathbf{U}\}$, $\mathbf{U}_q = \mathbf{Im}\{\mathbf{U}\}$ for

$$\mathbf{U} = \begin{bmatrix} u(M-1) & u(M-2) & \cdots & u(0) \\ u(M) & u(M-1) & \cdots & u(1) \\ \vdots & \vdots & \ddots & \vdots \\ u(N-1) & u(N-2) & \cdots & u(N-M) \end{bmatrix}$$

The predistorter $\mathbf{p}_i$ and $\mathbf{p}_q$ can be identified directly, without knowing or identifying IQ imbalance model $h_{11}(n), h_{12}(n), h_{21}(n), h_{22}(n)$ – Since IQ predistorter is the inverse function of IQ imbalance model, in perfect conditions, $u(n)$ and $y(n)$ are the same except for some time delay. So by replacing $u(n)$ with $y(n)$, (3.2) becomes

$$\mathbf{x} = \mathbf{Y}_i \mathbf{p}_i + \mathbf{Y}_q \mathbf{p}_q = \begin{bmatrix} \mathbf{Y}_i & \mathbf{Y}_q \end{bmatrix} \begin{bmatrix} \mathbf{p}_i \\ \mathbf{p}_q \end{bmatrix} \quad (3.3)$$

By solving (3.3), TX IQ imbalance predistorter $\mathbf{p}_i$ and $\mathbf{p}_q$ are identified.

To solve (3.3), a sequence of data $x(n)$ and $y(n)$ need to be acquired. Here $x(n)$ is available as it is the digital signal that goes into the DAC. $y(n)$ can be acquired using the test bed shown in Fig. 3-1 with DUT removed. Either Channel A or Channel C can be used, and the acquired IF signal is further digitally processed following the procedure in Fig. 3-2 to recover $y(n)$, i.e., the envelope of the RF signal that suffers from TX IQ imbalance.

3.2.4 Digital Demodulation

The advantage of digital demodulation [136] is that there are no IQ imbalance impairments in demodulation. The drawback is that it requires very high sampling rate – normally four times of IF carrier frequency or higher. Luckily, a modern oscilloscope usually has a RF bandwidth of more than 1 GHz, and a maximum sampling rate of several giga samples per second (Gsps). The high sampling rate makes it easy to acquire the signal;

moreover, it increases the time resolution of signal time alignment to counter the channel delay effects.

The received IF digitalized signal can be denoted as:

$$\begin{aligned}\tilde{r}(nT_s) = & I'(nT_s)\cos(2\pi(f_0 + \Delta f)nT_s + \phi) \\ & - Q'(nT_s)\sin(2\pi(f_0 + \Delta f)nT_s + \phi) + N(nT_s)\end{aligned}\quad (3.4)$$

where f_s is the sampling frequency, $T_s = 1/f_s$ is the sample interval, f_0 is the nominal intermediate frequency, Δf is the frequency offset from f_0 due to the local oscillator (LO) frequency difference in the down converters, ϕ is a constant phase, and $N(nT_s)$ is the channel noise. The digital demodulation is shown in Fig. 3-5.

Fig. 3-5 Digital demodulation

It uses incoherent demodulation, so there is a phase offset θ in LO. The higher order components are filtered out by digital low pass filters (LPF), which are also down samplers that reduce signals to baseband sampling rate. Let f_{bb} denote the baseband

sampling rate, and $T_{bb}=1/f_{bb}$ is the baseband sampling interval. With perfect time alignment and down sampling operation, the baseband signal is

$$\begin{aligned} r(mT_{bb}) &= \tilde{r}_l(mT_{bb})e^{j(2\pi\Delta f mT_{bb}+\theta)} \\ &= [I'(mT_{bb}) + j \cdot Q'(mT_{bb})]e^{j(2\pi\Delta f mT_{bb}+\theta)} \end{aligned} \quad (3.5)$$

where $\tilde{r}_l(mT_{bb})$ is the low pass equivalent of the received IF signal. Here it is shown that there are both frequency offset Δf and phase offset θ impairments in the received signals.

3.2.5 Efficient Implementation of Down Sampling and LPF

As shown in Fig. 3-5, the LPF accomplishes three goals:

- 1) It filters out the high frequency components.
- 2) It reduces the sampling rate to baseband sampling rate, based on which the PA behavior is characterized and modeled. The down sample ratio is $D = f_s / f_{bb}$.
- 3) It completes the first step of time alignment by finding the optimal down sample position out of every D high-rate samples.

The first two goals and the corresponding designs are introduced and analyzed in this subsection, while the third one will be analyzed in the next subsection.

Fig. 3-6 General block diagram of down sampler

Fig. 3-6 shows the general block diagram of the down sampler. The input signal $x(n)$ is convoluted with an anti-aliasing filter $h(n)$, and then down sampled by a factor D . The down sampled output $y(m)$ is given by:

$$y(m) = z(mD) = \sum_{k=0}^N h(k)x(mD-k) \quad (3.6)$$

Fig. 3-7 Direct-Form FIR filter structure with embedded down sampler

However, the above structure is very inefficient, because the filter is operating in high sampling rate, whereas only one out of every D filtered sampled is a valid output. In other words, most of the filtering computation is wasted. To improve the efficiency, the

down sampling operation can be embedded into the filter. The direct-form FIR filter structure [137] is shown in Fig. 3-7. Now all the multiplications and additions are calculated in baseband sampling rate, which reduces the computational complexity by a factor D .

The computation efficiency can be further improved by reducing the length of the FIR, from one M -tap filter into a set of smaller filters of length $k = M / D$, where M is a multiple of D . The structure is shown in Fig. 3-8.

Fig. 3-8 Polyphase filter structure

It is called *Polyphase* filter [137], which can be seen as a set of subfilters. Each time, only one subfilter of length $K = M / D$ is calculated. Hence the efficiency is further increased by another factor D . The subfilters are defined as

$$Q_l(n) = h(nD - l), \quad l = 0, 1, \dots, D-1 \quad (3.7)$$

3.2.6 Time Alignment

3.2.6.1 Channel Delay Difference

As shown in Fig. 3-1, the PA's input and output signals pass through Channel A and Channel C respectively and their baseband signals are used to characterize the PA behavior. Since time delays through Channel A and Channel C are different, Sig A and Sig C are misaligned in time. This misalignment [138] could result in degradation of the accuracy of signal envelop recovery and the behavioral model. Hence the time misalignment should be compensated before using the recovered envelopes to characterize the behavior of the power amplifier.

3.2.6.2 Training Sequences

To counter channel delay, frequency offset and phase offset impairments, training sequences are transmitted. Here 10 short training sequences and 2 long training sequences, both defined in IEEE802.11a [5], are transmitted, followed by data payload that is used to characterize PA behavior. The signal structure is shown in Fig. 3-9.

10 short training sequences	2 long training sequences	Data payload characterizing PA behavior
-----------------------------	---------------------------	---

Fig. 3-9 Signal structure

3.2.6.3 Time Alignment method

Instead of finding out the channel delay difference between Channel A and Channel C and then compensating the difference, both received signals, Sig A and Sig C, are aligned with the receiver timer. Based on the signal structure shown in Fig. 3-9, the first sample in

the short training sequences marks the start of the signals. If the start samples of both Sig A and Sig C can be precisely found, the start samples can be used as the time origins, from which the receiver starts to receive and process the two channel signals. In this way, Sig A and Sig C are aligned in time with receiver timer, and hence are aligned to each other.

Now the task is to precisely find the start sample of each signal. This is often called *Time Synchronization* in OFDM systems. There is a lot of research work on how to solve *Time Synchronization* problem [139-141]. A classical method is to do cross correlation between received short training sequences and the ideal ones – distortion-free version of the short training sequences stored in the receiver [142]. Depending on the number of ideal training sequences used, the magnitude of cross correlation results has one or more peaks, the first of which marks the start time of the received signal. Let $v(n)$ denote the ideal short training sequences, $r(n)$ denote the received signal, p denote the first sample of the signal or the first peak position, and R_{xx} denote the normalized cross correlation.

$$R_{xx}(m) = \frac{\sum_{n=0}^L v^*(n)r(n+m)}{\sum_{n=0}^L |v^*(n)|^2} \quad (3.8)$$

where L is length of ideal short training sequences, and

$$\begin{aligned} p &= \arg \max_m |R_{xx}(m)| \\ &= \arg \max_m \left| \frac{\sum_{n=0}^L v^*(n)r(n+m)}{\sum_{n=0}^L |v^*(n)|^2} \right| \end{aligned} \quad (3.9)$$

Here a modification is made on the classical method – instead of being computed in baseband sampling rate after the down sampler, here cross correlation is done in high

sampling rate before down sampling operation, in order to increase the resolution. An up sampling operation by interpolation may even be needed to increase time synchronization resolution, which in turn increases the accuracy of signal recovery and PA behavioral characterization and modeling. The cost is more computation resources.

3.2.6.4 Two Steps of Time Alignment

To implement time alignment scheme, two steps are designed:

The first step is to find the optimal down sampling position – which sample out of every D samples to be chosen to output in the down sampler. It is done in high sampling rate, and determines the resolution of time alignment.

The second step is to find the start sample in baseband sampling rate. This is the same as the classical *Time Synchronization* method.

In each step, a cross correlation between the received short training sequences and the ideal ones is calculated – the first is in high sampling rate, while the second is in baseband sampling rate. For the first step, the first cross correlation peak determines the optimal down sampling position, or which sample out of every D samples to output in the down sampler. Let p denote the first correlation peak position, and d denote the optimal down sampling position, then

$$d = p \bmod D \quad (3.10)$$

For implementation simplicity, the down sampler rate is fixed, while a buffer of adjustable length is inserted before the down sampler to facilitate changing down sampling position. The final diagram of demodulation with LPFs, down samplers, and the first step of time alignment is shown in Fig. 3-10.

Fig. 3-10 Final digital demodulation with LPFs, down samplers, and the first step of time alignment – finding optimal down sampling position.

3.2.7 Frequency Offset Compensation

Due to the inaccuracy of LOs in the up converters and down converters, frequency offset is a typical channel impairment which can severely distort signals and degrade the accuracy of envelope recovery for the PA behavioral characterization and modeling. The effect of frequency offset on OFDM signals is assessed in terms of SNR loss in [140], given by

$$\begin{aligned}
 \gamma(\Delta f) &= \frac{SNR^{(ideal)}}{SNR^{(real)}} \\
 &\approx 1 + \frac{1}{3} SNR^{(ideal)} \cdot (\pi N T_{bb} \Delta f)^2 \\
 &= 1 + \frac{1}{3} \frac{E_s}{N_0} \cdot (\pi \Delta \epsilon)^2
 \end{aligned} \tag{3.11}$$

where N is FFT size, $SNR^{(ideal)}$ is the SNR without frequency offset, $SNR^{(real)}$ is the SNR with frequency offset Δf , and $\Delta \epsilon = N T_{bb} \Delta f$ is the normalized frequency offset.

The algorithm proposed in [143] is used to estimate the frequency offset Δf , and then compensate it. For $2H$ repetitive training sequences, each having a length of P samples, the normalized frequency offset is estimated as

$$\Delta \hat{\varepsilon} = \frac{H}{\pi} \sum_{l=1}^H \omega(l) \arg \{ R(l) R^*(l-1) \} \quad (3.12)$$

where the weights are

$$\omega(l) = 3 \cdot \frac{(2H-l)(2H-l+1) - H^2}{H(4H^2 - 1)} \quad (3.13)$$

and the qP -lag autocorrelations are

$$R(l) = \sum_{k=1}^{N-lP} r(k+lP) r^*(k) \quad (3.14)$$

In the test bed, two steps are designed to estimate the frequency offset.

1) In the first step, 10 short training sequences are used, where $H=5$. By averaging over 5 weighted estimated values, the noise influence is greatly reduced. However, due to their short length, short training sequences cannot estimate small frequency offset, or the estimate error is relatively large. So this step is called coarse frequency offset estimation.

2) In the second step, 2 long training sequences are used, where $H=1$. Since their length is 5 times longer than that of the short training sequences, the estimation resolution is also increased by a factor 5. Hence it is called fine frequency offset estimation.

3.2.8 Phase Offset Compensation

After compensating the frequency offset, the received signals still suffer from phase offset impairments. Phase offset is a constant phase shift added to the recovered envelopes

due to inherent demodulation. Moreover, since phase offsets are different in different experiments, the PA AM/PM characterization suffers random phase shift or phase uncertainty over different experiments. Hence phase offset should be compensated before the envelopes of the PA's input and output signals can be accurately recovered.

Similar to the time alignment, the method to estimate phase offset is based on the cross correlation between the received long training sequences and the ideal ones – distortion-free version of the long training sequences stored in the receiver [144]. Let $u(n)$ denote the ideal long training sequences, and $w(n)$ denote the received long training sequences suffering from phase offset θ . Without considering noise, and assuming perfect time alignment and frequency offset compensation, it is in the form:

$$w(n) = u(n)e^{j\theta} \quad (3.15)$$

Then the normalized cross correlation R_{xx} is

$$R_{xx} = \frac{\sum_{n=0}^K u^*(n)w(n)}{\sum_{n=0}^K |u^*(n)|^2} = e^{j\theta} \quad (3.16)$$

where K is the length of the long training sequences. Hence the phase offset estimation is the angular of the cross correlation results

$$\begin{aligned} \hat{\theta} &= \text{angular}\{R_{xx}\} \\ &= \text{angular} \arg \left\{ \frac{\sum_{n=0}^K u^*(n)w(n)}{\sum_{n=0}^K |u^*(n)|^2} \right\} \end{aligned} \quad (3.17)$$

3.3 TX IQ Imbalance Test Bed

With some modification, the proposed PA test bed can be used to characterize TX IQ imbalance, and test the performance of their behavioral models. Three changes need to be done:

1. Remove the power amplifier from the original test bed (in Fig. 3-1).
2. Remove the TX IQ imbalance predistortion from the baseband signal processing part at the transmitter, because the TX IQ imbalance is to be characterized by behavioral models. The baseband part of the new transmitter is shown in Fig. 3-11.
3. Only one down-converter channel (Channel A or Channel C) in the original test bed (in Fig. 3-1) is used to down-convert the signal from RF to IF, while the other is not connected.

Fig. 3-11 Transmitter digital signal processing for TX IQ imbalance Test Bed

The overall setup of the TX IQ imbalance test bed is shown in Fig. 3-12. The TX IQ imbalance inside the signal generator E4438C is to be characterized. Note that unlike the PA test bed, the input of the behavioral model of the TX IQ imbalance is the digital

baseband transmitting signal that is saved in the computer. Also notice that an additional IQ imbalance can be manually added in the baseband of the transmitter to simulate a low cost modulator having poor IQ imbalance performance.

Fig. 3-12 Test Bed for TX IQ imbalance. The dotted block represents an additional IQ imbalance that can be manually added in the transmitter.

3.4 Joint TX IQ Imbalance and PA Test Bed

The proposed PA test bed can also be modified in a way that the test bed can model the joint TX IQ imbalance and PA distortions. Similar to the TX IQ imbalance test bed in Section 3.3, the TX IQ imbalance predistortion should be removed from the transmitter, and only one down-converter channel is used. The input of the behavioral model is also the digital baseband transmitting signal that is saved in the computer. And any additional IQ imbalance can also be conveniently added in the baseband transmitter, providing a good flexibility. The overall setup is shown in Fig. 3-13.

Fig. 3-13 Test Bed for joint TX IQ Imbalance and PA

3.5 Performance Metrics Design

There are generally two goals in using the test bed: The first goal is to recover the envelopes of the DUT's input and output signals in the form of digital samples, which are later used to characterize and model the DUT behavior. The second goal is to evaluate the performance of DUT behavioral modeling and predistortion schemes. For the former, competent metrics are required to evaluate how accurate the envelopes can be recovered; while for the latter, metrics are needed to judge the wellness of the DUT behavioral models.

3.5.1 Recovery Accuracy Metric

The difficulty in evaluating the accuracy of signal recovery is that the ideal envelopes of the DUT's input and output signals cannot be directly acquired to compare with the

recovered ones. Here a new way to evaluate the recovery accuracy is introduced to overcome this problem. The basic idea is to use the transmitter baseband digital waveform to approximate the complex envelope of the RF signal when the DUT is removed.

Assume that there are no distortions or impairments in the signal generator (Agilent E4438C ESG). In another word, assume the DAC and up-conversion is absolutely linear. Then the transmitter baseband digital waveform just before the DAC can be regarded as the low-pass equivalent or the complex envelope of the RF signal at point A (or C) point in Fig. 3-1. Hence the difference between the transmitter baseband digital waveform and the recovered envelope (through Channel A or Channel C) is a competent metric to evaluate the recovery accuracy. It is defined as the relative envelope error (REE) given by:

$$REE = \sqrt{\frac{\sum_{n=1}^N |s(n) - r(n)|^2}{\sum_{n=1}^N |s(n)|^2}} \quad (3.18)$$

where $s(n)$ and $r(n)$ are the transmitter baseband digital waveform and the recovered envelope, respectively. If the REE values of both Channel A and Channel C are very small, the test bed is regarded as accurately recovering the envelopes of the DUT's input and output signals.

3.5.2 EVM

Error vector magnitude (EVM) is another convenient way to evaluate the linearity of the channels; moreover, it can also evaluate the performance of the DUT predistortion methods using behavioral models. EVM is the difference between the received constellation and the transmitted one. It is sensitive to most kinds of impairments,

including nonlinear distortion. Hence a small value of EVM means a good linearity of the test bed and its competency to accurately recover the envelope of the DUT's input and output signals. EVM is defines as [145]

$$EVM = \frac{\sum_{n=1}^N |X(n) - Y(n)|^2}{\sum_{n=1}^N |X(n)|^2} \quad (3.19)$$

where $X(n)$ and $Y(n)$ are the transmitted constellation and received constellation, respectively. The difference between (3.18) and (3.19) can be clearly illustrated in OFDM signals. In (3.18), $s(n)$ and $r(n)$ are both samples in the time domain; while in (3.19), $X(n)$ and $Y(n)$ are both constellations in the frequency domain, whose sampling rate is usually lower than $s(n)$ and $r(n)$.

3.5.3 Spectrum Estimation and ACPR

Spectral power density (PSD) and adjacent channel power ratio (ACPR) are figure-of-merits to assess the DUT nonlinear and memory effects, and can also be used to judge the wellness of DUT behavioral modeling and predistortion schemes.

Usually a spectrum analyzer is used to measure signal spectrum, but it cannot be used in the simulation that evaluates the DUT behavioral modeling. Instead, numerical spectrum estimation can be used both in real experiment and in simulation. The traditional numerical spectrum estimation uses periodogram [146], which uses the amplitudes of signal FFT transformation as the spectrum. But periodogram is not a consistent estimation of the true spectrum; in another word, its variance remains high [137]. In our test bed, Welch method [147] is applied to estimate signal spectrum, which is justified to be a

consistent estimation. Welch method divides the signal into several segments, calculates the periodogram on each segment, and averages them as the final spectrum. Moreover, by overlapping between adjacent segments and using proper time window, Welch method reduces the required length of data to achieve sufficient accuracy.

Fig. 3-14 Welch method: segmentation with 50% overlap

Fig. 3-14 shows the segmentation in Welch estimation with 50% overlap ratio. Assume the total number of segment is Q , and the length of each segment is P . Each segment of data is denoted by $x^{(q)}(n)$, and the time window is denoted by $w(n)$. For each segment, the periodogram is

$$S_{est}^{(q)}(k) = \frac{1}{PU} \left| \sum_{n=0}^{P-1} w(n) x^{(q)}(n) e^{-j2\pi kn/P} \right|^2 \quad (3.20)$$

where the normalization factor U is

$$U = \frac{1}{P} \sum_{n=0}^{P-1} w^2(n) \quad (3.21)$$

Then the final Welch spectrum is the average of $S_{est}^{(q)}(k)$ as

$$S_{est}(k) = \frac{1}{Q} \sum_{q=1}^Q S_{est}^{(q)}(k) \quad (3.22)$$

After estimating the signal spectrum, Adjacent Channel Power Ratio (ACPR) is calculated. Due to memory effects, left side ACPR and right side ACPR are different. The left side ACPR and right side ACPR are defined as [12]:

$$ACPR_{left} = \frac{\int_{f_c-f_o+BW/2}^{f_c-f_o-BW/2} |H(f)|^2 S(f) df}{\int_{f_L}^{f_U} |H(f)|^2 S(f) df} \quad (3.23)$$

and

$$ACPR_{right} = \frac{\int_{f_c+f_o+BW/2}^{f_c+f_o-BW/2} |H(f)|^2 S(f) df}{\int_{f_L}^{f_U} |H(f)|^2 S(f) df} \quad (3.24)$$

3.6 Hardware Design Consideration

3.6.1 Noise Issue

Reducing noise or increasing SNR is critical to enhance the accuracy of signal recovery. Generally speaking, there are three noise sources in the proposed test bed: noise from the signal generator, noise added in the channels of the test bed, and noise generated by the oscilloscope.

The first noise source can be reduced or limited to a very low level by using competent signal generators. Agilent E4438C ESG contains I/Q-channel DACs of 16-bit resolution, an inner synthesizer and inner mixers that can provide up to 3GHz carrier frequency. From experimental measurement, Agilent E4438C ESG can generate a SNR higher than 60 dB for OFDM signals, which means the noise in the signal source can be neglected.

Fig. 3-15 Down Converter Channels

The noise in the channels of the test bed can be greatly limited by carefully using amplifiers and filters in circuit design. In Fig. 3-15, the LO is implemented by Agilent 83623A Synthesizer with the output power level of 15 dBm. And the mixers are implemented by Mini-Circuits SIM-83+, whose conversion loss is 6 dB, LO-RF isolation is 35 dB, LO-IF isolation is 23 dB. The low pass filter (LPF) is succeeded by a low noise amplifier (LNA), which can greatly reduce the out-of-band noise level. On the other hand, the in-band noise level, determined by both the signal generator and the devices in channels, is very difficult to reduce. In our case, without the DUT in Fig. 3-1, the signal SNR at the input of oscilloscope can achieve as high as above 55 dB.

The quantization error in the oscilloscope is another mayor noise source. The SNR after quantization is determined by the input noise level and the dynamic range of the ADC. In our case, Infiniium 54832D MSO oscilloscope is used, which has 8-bit resolution, or vertical resolution is 0.4%. In this case, the maximum quantization SNR (MQSNR) is

around 47 dB. Usually the real SNR is much less than MQSNR because it is hard to let the signal perfectly fit the dynamic range of the ADC. Luckily, the SNR can be further improved later by using digital filters. In our case, without the DUT in Fig. 3-1, the final SNR after digital signal processing can achieve as high as 45 dB.

3.6.2 Linearity of Test Bed

Fig. 3-16 Calibration of the nonlinearity of down-converters: Gain versus input power level for Channel A and Channel C.

Besides SNR, the linearity of the test bed is another critical issue to determine the overall performance. Ideally speaking, the whole test bed without DUT should be absolutely linear. However, the DAC and ADC, the mixers in up-converter and down-converters, and the low noise amplifiers give rise to nonlinear effects of the test bed. Hence measures are required to limit the nonlinearity of the test bed. Here the main

nonlinear effects are generated from the down-converters as in Fig. 3-15, which include mixers and LNAs. In this case, careful calibration is required to find out the linear working zone of the down-converters, and then a proper back-off from the 1-dB compression point is chosen to ensure that the signal can pass linearly through the down-converters, while the signal power level is as high as possible to maintain high SNR value. The back-off value can be chosen slightly higher than the value of PAPR of the input signal.

Fig. 3-16 shows the calibration of the nonlinear characteristics of the two down conversion channels, where the measured 1-dB compression points of the input for both channels are above -15 dBm. The calibration is carried out by a single-tone signal. Please note that the output power level of DUT (a power amplifier) is usually much higher than -15 dBm (e.g. the output power level of ZVE-8G power amplifier is above 30 dBm). Hence to make sure the input power level is within the linear zone, a proper attenuator should be inserted before the down-converter, as shown in Fig. 3-1. For the case that a 16.6MHz OFDM signal with more than 10 dB PAPR is applied, the input of the down-converter is set at 15 dB back-off, or -30 dBm power level.

3.7 Performance Assessment

Experiments are executed to assess the performance of the PA test bed, i.e., its accuracy to recover the envelopes of the PA's input and output signals in the form of digital samples. Note that due to their similarity, the TX IQ imbalance test bed, and joint TX IQ Imbalance and PA test bed are not assessed here. For more information on their performance, please refer to the measurement results in Chapter 5 and Chapter 6.

3.7.1 Measurement Setup

The structure is shown in Fig. 3-1. Agilent E4438C ESG vector signal generator with output frequency ranging from 250KHz to 3GHz is used to generate arbitrary waveform signals, where modulation bandwidth is 100MHz. A 4-channel oscilloscope named Infiniium 54832D MSO with 1GHz analog bandwidth, 4GS/s maximum sample rate and 8-bit vertical resolution is used. The DUT is a Class-A Mini-Circuits ZVE-8G high power amplifier (HPA) with working frequency of 2.4GHz. The specifications and datasheet of ZVE-8G are shown in Table 3-1 and Table 3-2. The frequency characteristics of ZVE-8G over the frequency range 2GHz ~ 8GHz are shown in Fig. 3-17 and Fig. 3-18. Meanwhile, the frequency offset in the test bed is around 24 KHz, and the value of phase offset is uniformly distributed in the range $[0^\circ \ 360^\circ]$. These two impairments can be compensated by the methods in Section 3.2.7 and Section 3.2.8.

To assess the test bed, signals with training sequences as shown in Fig. 3-9 are generated with output frequency of 2.4GHz, and then passed through the DUT. The input and output signals of the DUT are down converted separately to IF of 50MHz, and received into the oscilloscope with a sampling rate of 2GS/s. The signal generator and oscilloscope are both connected to a PC through GPIB. Advanced Design System (ADS) and Matlab are cooperatively programmed to implement digital signal processing, PA behavioral modeling and metrics calculation. For instance, an OFDM baseband waveform is firstly generated in ADS, and then downloaded into E4438C through GPIB to generate the RF signal. The received signals in oscilloscope are transferred to PC through GPIB for further processing.

Frequency Range (MHz)	2000 ~ 8000
Input no damage (dBm)	20
IP3 (dBm) typ.	40
DC Voltage (V) Norm.	12
DC Current (A) max.	1.2
Operating temperature (°C)	-55 ~ 54

Table 3-1 ZVE-8G specifications

Frequency (MHz)	Gain (dB)	Noise Figure (dB)	Pout at 1 dB compr. (dBm)
2000	35.04	5.23	31.11
2300	35.84	4.89	32.06
2600	36.02	4.44	32.27
3200	35.54	4.16	32.63
3500	35.33	4.11	32.74
3800	35.25	4.05	32.72
4100	35.3	4.05	32.13
4400	35.08	4.09	32.36
4700	34.96	4.1	32.37
5000	34.86	4.08	32.56
5300	34.78	4.15	31.74
5600	34.74	4.16	32.04
5900	34.85	4.23	32.01
6200	35.03	4.37	32.08
6500	35.07	4.33	31.98
6800	35.37	4.26	31.97
7100	35.63	4.34	31.5
7400	35.95	4.36	31.59
7700	35.84	4.22	31.48
8000	35.41	4.2	31.57

Table 3-2 ZVE-8G datasheet

Fig. 3-17 ZVE-8G gain (dB) versus frequency

Fig. 3-18 Output 1dB compression point power versus frequency

3.7.2 Measurement Results on Envelope Recovery

3.7.2.1 EVM performance and IQ imbalance

The accuracy of the recovered envelope is dependent on the linearity of the channels. After removing the DUT in the test bed, EVM measurement can be used to assess the linearity of the channels and how well all kinds of impairments are compensated.

Fig. 3-19 and Fig. 3-20 show how IQ imbalance influences EVM measurement. The IQ imbalance is generated from the modulators, LOs, mixers, and reconstruction filters, etc. Both figures show the constellations of 16QAM plus BPSK signal after passing through the down conversion Channel A. The SNR is around 48dB. In Fig. 3-19, due to IQ imbalance in the E4438C ESG, the signal is distorted by mirror components and EVM is measured to be -34.3dB . After compensating IQ imbalance using the linear predistorter structure as shown in Fig. 3-4, the constellation is significantly improved as illustrated in Fig. 3-20, and EVM is measured to be -42.3dB . The IQ imbalance predistorter is in the form of (3.2), where the length M of each predistortion filter is 6 taps and the total number of coefficients is 24. The IQ imbalance predistorter is identified by solving (3.3).

Fig. 3-19 Constellation: 16QAM+BPSK, without compensating IQ imbalance in modulator, EVM is -34.3dB .

Fig. 3-20 Constellation: 16 QAM+BPSK, with compensated IQ imbalance in modulator, EVM is -42.3dB .

Channel Name	EVM value
Channel A	-42.3dB
Channel C	-41.3dB

Table 3-3 EVM measurement results for Channel A and Channel C

With IQ imbalance compensation, the EVM measurement results for Channel A and Channel C are shown in Table 3-3. Both EVM values are below -41dB, which justifies that the channels are quite linear and all the impairments can be compensated very well, and the test bed is able to accurately recover the envelopes of PA's input and output signals.

3.7.2.2 REE performance:

REE defined in (3.18) is measured to directly assess the accuracy of envelope recovery. Table 3-4 and Table 3-5 show the REE measurement results under different baseband sampling rate. A range of baseband sampling rate is required to satisfy PAs of different memory effects and signals of different bandwidth. From the measured REE values, the baseband sampling rate does not influence the accuracy of the envelope recovery very much. All the relative envelope errors between transmitter baseband digital waveform and the recovered envelopes are below 5.5%, representing a good envelope recovery.

Baseband sampling rate	REE value
20MS/s	5.25%

40MS/s	4.95%
80MS/s	4.79%
100MS/s	5.19%
200MS/s	4.90%

Table 3-4 Channel A: REE values versus baseband sampling rate

Baseband sampling rate	REE value
20MS/s	5.37%
40MS/s	4.57%
80MS/s	4.42%
100MS/s	5.43%
200MS/s	4.79%

Table 3-5 Channel C: REE values versus baseband sampling rate

Table 3-6 shows the REE measurement results under different resolution of time alignment by interpolation. An interpolation can be inserted before time alignment and down sampler in Fig. 3-10 to increase the resolution of time alignment. From the results,

with 10 times interpolation the REE performance is improved by around 1%, otherwise there is no significant difference.

Time resolution	Interpolation ratio	REE value
0.5 ns	1	4.90%
0.25 ns	2	5.01%
0.1667 ns	3	4.73%
0.125 ns	4	4.68%
0.1 ns	5	5.01%
0.05 ns	10	3.94%

Table 3-6 Channel A: REE value versus time alignment resolution

Fig. 3-21 and Fig. 3-22 show the recovered envelopes of the RF signal versus the transmitted baseband waveform, with the DUT removed. The baseband sampling rate at the receiver is 100MS/s, and the resolution of time alignment is 0.5ns. Ideally the test bed is perfectly linear, and the two waveforms should be the same. Although distorted by noise and digital quantization error, the real part of recovered envelope is still approximately coincided with that of the transmitted baseband waveform, justifying the competency of the test bed to accurately recover the envelopes of the PA's input and output signals.

Fig. 3-21 Time-domain waveforms comparison: Removing DUT, the real part of recovered envelope waveform by Channel C versus the real part of transmitted baseband waveform.

Fig. 3-22 Time-domain waveforms comparison: Removing DUT, the image part of recovered envelope waveform by Channel C versus the image part of transmitted baseband waveform.

Fig. 3-23 Recovered AM/AM behavior of ZVE-8G HPA: The curve is drawn from the recovered envelopes of PA's input and output signals.

Fig. 3-24 Recovered AM/PM behavior of ZVE-8G HPA: The curve is drawn from the recovered envelopes of PA's input and output signals.

Finally, in Fig. 3-23 and Fig. 3-24, the behavior of the DUT is characterized from the recovered envelopes of the PA's input and output signals – the AM/AM and AM/PM figures. Due to the memory effects and the signal of 16.6MHz bandwidth, the AM/AM and AM/PM are not curves but scattered bunches of points. The 1-dB gain compression input power point is around 0 dBm, after which, the output power quickly saturates. The test bed can accurately recover the envelopes for the PA behavioral characterization and modeling.

3.8 Conclusion

A new test bed is proposed for TX IQ Imbalance and PA behavioral modeling and characterization. It generally accomplishes two goals: The first goal is to accurately recover the envelopes of the DUT input and output signals to facilitate behavioral characterization and modeling. The second goal is to provide necessary metrics to evaluate the accuracy of envelope recovery, and the performance of behavioral model and predistortion.

The proposed test bed is based on general purpose lab equipment with software running on a PC. This opens up flexibility for changes to the test bed hardware and software to facilitate different test scenarios or for enhanced performances. The proposed test bed can generate signals of arbitrary waveform with enough linear power to drive the DUT. Software signal processing methods are implemented to compensate the major impairments, such as channel delay, frequency offset, and carrier phase offset. The hardware is designed to provide channels of both high SNR and high linearity.

The test bed also provides the facilities and metrics to evaluate the accuracy of envelope recovery, and the performance of behavioral model and predistortion scheme. Relative envelope error (REE) is proposed as a metric to evaluate the accuracy of envelope recovery. Signal spectrum is estimated by Welch method with arbitrary accuracy, and consequently ACPR values can be calculated. Constellation and EVM value are also used to evaluate the linearity of the channel, the accuracy of a behavioral model, and the wellness of a linearization scheme.

The competence of the test bed is demonstrated by a number of experiment results. It achieves high accuracy for envelope recovery under different baseband sampling rates or different interpolation ratios in time alignment scheme. The AM/AM and AM/PM behaviors are characterized based on the recovered envelopes. Further experiment results of behavioral models on TX IQ imbalance and power amplifier are shown in Chapter 5 and Chapter 6.

Chapter 4 A New Power Amplifier Predistortion Architecture Based on Memory Polynomial Model

4.1 Introduction

Digital predistortion is a cost-effective method to linearize the power amplifier. Polynomial based models have long been researched [148, 149]. In recent years, in order to predistort the memory effects of the power amplifier, a modified polynomial model – memory polynomial model is introduced [41]. It is in fact a simplified version of the Volterra Model, and demonstrated to be efficient to linearize PA and reduce ACPR [58]. In [34], the memory polynomial model has been researched more generally, introducing two sets of cross-terms.

However, most research on memory polynomial model is based on a single architecture – predistorting the amplitude and phase of the signal simultaneously. On the contrary, power amplifiers have two characteristics – AM/AM and AM/PM relations. Even if they are not independent to each other, AM/AM and AM/PM comply with different relations. From this point of view, predistorting amplitude and phase separately can improve the effects of predistortion, and get a more linearized signal at the output side of the PA. To be more precise, the new architecture can improve the performances of ACPR and EVM.

Earlier work employing a model of predistorting AM/AM and AM/PM separately with polynomials was report in [148, 149], but the work did not include memory effects.

In this chapter, a new memory-polynomial-model-based architecture predistorting the amplitude and phase separately is proposed. According to the characteristics of AM/AM and AM/PM relations, the memory polynomial models in the new architecture should be real-valued. Simulation results show significant improvements in ACPR and EVM, compared with the traditional architecture.

The organization of the chapter is as follows. In Section 4.2, the AM/AM and AM/PM functions of a power amplifier with significant memory effects are shown, and then real-valued memory polynomial model is deducted. In Section 4.3, the new predistortion architecture using memory polynomial model is proposed, its mathematical analysis is presented, and its predistortion block diagram is illustrated. And in Section 4.4, the simulation results are shown, and compared with those of the traditional architecture. Finally, Section 4.5 gives the conclusion.

4.2 Memory Effects and Memory Polynomial Model

4.2.1 Memory Effects on AM/AM and AM/PM Relations of PA

The first step to predistort and linearize a PA is to model its behavioral characteristics accurately. Previously, extensive research has been done on modeling the nonlinearity of PA based on its AM/AM and AM/PM relations without memory. However, in recent years more and more wireless systems use wide-band signals such as OFDM and CDMA. The memory effects of the PA in such systems cannot be ignored and contribute significantly to

the distortion of the signal. These memory effects arise from both the circuit and the thermal effects of the transistor.

Memory effects change the AM/AM and AM/PM relations of the PA. The relations do not only depend on the current input signal, but also on the history of the past input signal – a sequence of past inputs. As a result, with different sequence of past inputs, a current input can be mapped to many different outputs. The AM/AM and AM/PM become one-to-many mapping. Fig. 2-1 and Fig. 2-2 show a typical AM/AM and AM/PM relations, where samples are scattering within certain ranges.

Treating amplitude and phase models separately has been applied to characterize the behavior of the power amplifier. In [150], AM/AM and AM/PM relations are modeled separately by polynomial based models. A modified Saleh model is proposed to characterize the AM/AM and AM/PM separately in [151]. Polynomial based models have also been applied to predistort AM/AM and AM/PM separately [148, 149]. However, these models only characterize the nonlinear memoryless behavior of the power amplifier. Then in [152], Hammerstein model and neural network are combined together to characterize the nonlinear memory behavior of the power amplifier by modeling AM/AM and AM/PM relations separately. However, the structure is complex and difficult to identify.

For any given power amplifier (PA), the AM/AM and AM/PM relations can be expressed as two different functions:

$$|y(n)| = f(|x(n)|, |x(n-1)|, \dots, |x(n-m_{am}^0)|) \quad (4.1)$$

$$\angle y(n) = g(|x(n)|, |x(n-1)|, \dots, |x(n-m_{pm}^0)|) \quad (4.2)$$

Thus the natural inverse of the PA is to inverse AM/AM and AM/PM relations separately:

$$|x(n)| = f'(|y(n)|, |y(n-1)|, \dots, |y(n-m_{am})|) \quad (4.3)$$

$$\angle x(n) = g'(|y(n)|, |y(n-1)|, \dots, |y(n-m_{pm})|) \quad (4.4)$$

Equation (4.3) and (4.4) predistort the amplitude and phase of the signal separately, and together compose the inverse function of the PA that can be a good digital predistorter.

4.2.2 Volterra vs. Memory Polynomial Model

The traditional predistortion model is the Volterra model. And there is a technique known as *pth-order inverse* [46] that can be used to derive its inversed model. Thus, by modeling the PA's behavior, its inversed model can be directly derived and used as the predistorter. The discrete time form of Volterra model with *K-order* and *M-tap* memory can be written as [46]:

$$y(n) = \sum_{k=1}^K y_k(n) \quad (4.5)$$

where

$$y_k(n) = \sum_{m_1=0}^{M-1} \dots \sum_{m_k=0}^{M-1} h_k(m_1, \dots, m_k) \prod_{l=1}^k x(n-m_l) \quad (4.6)$$

However, Volterra model is too complex and has too many parameters, thus it cannot be used in a practical system.

Many methods have been used to simplify Volterra model. In [41], it only uses the diagonal parameters and assumes $m_1 = m_2 = \dots = m_k = m$, and (4.5) becomes:

$$y_k(n) = \sum_{m=0}^{M-1} h_k(m) \cdot x^k(n-m) \quad (4.7)$$

Moreover, the signal's bandwidth is a small percentage of the carrier frequency and the components of interest are close to the carrier frequency. Thus, (4.7) can be further simplified [34]:

$$y(n) = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} c_{km} x(n-m) |x(n-m)|^k \quad (4.8)$$

Equation (4.8) is the memory polynomial model (MPM), where $x(n)$ and $y(n)$ are complex-valued input and output signals respectively.

Under some conditions (i.e., image part of the signal is zero), $x(n)$ and $y(n)$ become real-value, then the memory polynomial model also becomes real-valued:

$$y(n) = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} c_{km} [x(n-m)]^{k+1} \quad (4.9)$$

where c_{km} is real-valued.

4.3 New DPD Architecture Based on MPM

4.3.1 New Architecture

Most predistortion schemes so far use only one MPM to predistort both the amplitude and phase of the input signal simultaneously. In this section, we explore the benefits of predistorting the amplitude and phase shift separately with two different memory polynomial models. The predistorter has the form as:

$$\begin{aligned} x(n) &= |x(n)| e^{j\angle x(n)} \\ &= |x(n)| e^{j(\angle u(n) + \Phi)} \end{aligned} \quad (4.10)$$

where

$$\begin{aligned}
|x(n)| &= A(|u(n)|, \dots, |u(n-M_1)|) \\
&= \sum_{k=0}^{K_1-1} \sum_{m=0}^{M_1-1} a_{km} |u(n-m)|^{k+1}
\end{aligned} \tag{4.11}$$

and

$$\begin{aligned}
\angle x(n) &= \angle u(n) + \Phi(|u(n)|, \dots, |u(n-M_2)|) \\
&= \angle u(n) + \sum_{k=0}^{K_2-1} \sum_{m=0}^{M_2-1} b_{km} |u(n-m)|^{k+1}
\end{aligned} \tag{4.12}$$

where $x(n)$ and $u(n)$ are the output and the input of the predistorter respectively. And $x(n)$ and $u(n)$ are both complex-valued signals. Note that the coefficients $\{a_{km}\}$ and $\{b_{km}\}$ are both real-valued. And the complex-valued (4.10) is converted to two real-valued MPM models expressed in (4.11) and (4.12), which follow the mathematical form as defined in (4.9).

The two MPMs use the same structure and identification methods, and only their coefficients are different. This ensures that the DPD can be conveniently identified and implemented.

4.3.2 Analysis of the New Architecture

A deeper mathematical analysis can reveal more of the new architecture, and its relationship to Volterra Model.

Equation (4.11) and (4.12) can be rewritten as:

$$\begin{aligned}
|x(n)| &= f(u(n), \dots, u(n-M_1)) \\
&= \sum_{k=0}^{K_1-1} \sum_{m=0}^{M_1-1} \tilde{a}_{km} u(n-m) |u(n-m)|^k
\end{aligned} \tag{4.13}$$

And

$$\begin{aligned}
\angle x(n) &= \angle u(n) + g(u(n), \dots, u(n-M_2)) \\
&= \angle u(n) + \sum_{k=0}^{K_2-1} \sum_{m=0}^{M_2-1} \tilde{b}_{km} u(n-m) |u(n-m)|^k
\end{aligned} \tag{4.14}$$

where the complex-valued equivalent coefficients are

$$\begin{cases} \tilde{a}_{km} = a_{km} \cdot \frac{|u(n-m)|}{u(n-m)} = a_{km} e^{-j\angle u(n-m)} \\ \tilde{b}_{km} = b_{km} \cdot \frac{|u(n-m)|}{u(n-m)} = b_{km} e^{-j\angle u(n-m)} \end{cases} \tag{4.15}$$

And K_1 and M_1 are the nonlinear order and memory depth of the amplitude function $f(n)$ respectively. K_2 and M_2 are the nonlinear order and memory depth of the phase shift function $g(n)$ respectively. Moreover, K_1 and K_2 are independent, M_1 and M_2 are independent.

The new DPD architecture of MPMs can be rewritten in the form:

$$\begin{aligned}
x(n) &= |x(n)| e^{j\angle x(n)} \\
&= f(u(n), \dots, u(n-M_1)) e^{jg(u(n), \dots, u(n-M_2))} e^{j\angle u(n)}
\end{aligned} \tag{4.16}$$

Expanding e^x by its Taylor Series:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \tag{4.17}$$

Now expand (4.16) by two steps: (1) Firstly, only consider the first two elements in (4.17) to simplify the analysis; (2) In the next step, include high-order items in (4.17).

Firstly, only consider the first two elements on the right hand of (4.17), and the new architecture becomes (ignore the constant $e^{j\angle u(n)}$):

$$\begin{aligned}
x(n) &= f(u)e^{j \cdot g(u)} \\
&= f(u)[1 + j \cdot g(u)] \\
&= f(u) + j \cdot f(u)g(u) \\
&= \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \tilde{a}_{k_1 m_1} u(n-m_1) |u(n-m_1)|^{k_1} \\
&\quad + \left(j \cdot \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \tilde{a}_{k_1 m_1} u(n-m_1) |u(n-m_1)|^{k_1} \right) \cdot \left(\sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \tilde{b}_{k_2 m_2} u(n-m_2) |u(n-m_2)|^{k_2} \right)
\end{aligned} \tag{4.18}$$

Considering the second term in (4.18), it can be decomposed as:

$$\begin{aligned}
&\left(j \cdot \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \tilde{a}_{k_1 m_1} u(n-m_1) |u(n-m_1)|^{k_1} \right) \cdot \left(\sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \tilde{b}_{k_2 m_2} u(n-m_2) |u(n-m_2)|^{k_2} \right) \\
&= j \cdot \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \tilde{a}_{k_1 m_1} \tilde{b}_{k_2 m_2} u(n-m_1) u(n-m_2) \cdot |u(n-m_1)|^{k_1} |u(n-m_2)|^{k_2}
\end{aligned} \tag{4.19}$$

Hence (4.18) becomes:

$$\begin{aligned}
x(n) &= \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \tilde{a}_{k_1 m_1} u(n-m_1) |u(n-m_1)|^{k_1} \\
&\quad + j \cdot \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \tilde{a}_{k_1 m_1} \tilde{b}_{k_2 m_2} u(n-m_1) u(n-m_2) \cdot |u(n-m_1)|^{k_1} |u(n-m_2)|^{k_2}
\end{aligned} \tag{4.20}$$

Equation (4.20) reveals two things for the new DPD architecture of MPMs. Firstly, it increases the highest nonlinear order of the DPD model, from $\max(K_1, K_2)$ to $K_1 + K_2$. Secondly, it introduces cross terms in memory, i.e., the 2nd-order cross term $u(n-m_1)u(n-m_2)$.

Now in the second step, consider the n^{th} order element $x^n/n!$ in (4.17), its multiplication with $f(u)$ is

$$\begin{aligned}
& f(u) \left(j \cdot g(u) \right)^n \\
&= \left(\frac{j}{n!} \right)^n \cdot \left(\sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \tilde{a}_{k_1 m_1} u(n-m_1) |u(n-m_1)|^{k_1} \right) \cdot \left(\sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \tilde{b}_{k_2 m_2} u(n-m_2) |u(n-m_2)|^{k_2} \right)^n \\
&= \left(\frac{j}{n!} \right)^n \cdot \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \cdots \sum_{k_{n+1}=0}^{K_{n+1}-1} \sum_{m_{n+1}=0}^{M_{n+1}-1} \tilde{a}_{k_1 m_1} \tilde{b}_{k_2 m_2} \cdots \tilde{b}_{k_{n+1} m_{n+1}} \\
&\quad \cdot u(n-m_1) u(n-m_2) \cdots u(n-m_{n+1}) |u(n-m_1)|^{k_1} |u(n-m_2)|^{k_2} \cdots |u(n-m_{n+1})|^{k_{n+1}} \\
&= \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \cdots \sum_{k_{n+1}=0}^{K_{n+1}-1} \sum_{m_{n+1}=0}^{M_{n+1}-1} \tilde{h}(k_1, m_1, k_2, m_2, \dots, k_{n+1}, m_{n+1}) \\
&\quad \cdot u(n-m_1) u(n-m_2) \cdots u(n-m_{n+1}) |u(n-m_1)|^{k_1} |u(n-m_2)|^{k_2} \cdots |u(n-m_{n+1})|^{k_{n+1}}
\end{aligned} \tag{4.21}$$

where the coefficient

$$\tilde{h}(k_1, m_1, k_2, m_2, \dots, k_{n+1}, m_{n+1}) = \left(\frac{j}{n!} \right)^n \sum_{m_{n+1}=0}^{M_{n+1}-1} \tilde{a}_{k_1 m_1} \tilde{b}_{k_2 m_2} \cdots \tilde{b}_{k_{n+1} m_{n+1}} \tag{4.22}$$

Altogether, the final mathematical expansion for the new DPD architecture of MPM can be written as

$$\begin{aligned}
x(n) &= \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \tilde{h}(k_1, m_1) u(n-m_1) |u(n-m_1)|^{k_1} \\
&+ j \cdot \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \tilde{h}(k_1, m_1, k_2, m_2) u(n-m_1) u(n-m_2) \cdot |u(n-m_1)|^{k_1} |u(n-m_2)|^{k_2} \\
&+ \cdots \\
&+ \sum_{k_1=0}^{K_1-1} \sum_{m_1=0}^{M_1-1} \sum_{k_2=0}^{K_2-1} \sum_{m_2=0}^{M_2-1} \cdots \sum_{k_{n+1}=0}^{K_{n+1}-1} \sum_{m_{n+1}=0}^{M_{n+1}-1} \tilde{h}(k_1, m_1, k_2, m_2, \dots, k_{n+1}, m_{n+1}) \\
&\quad \cdot u(n-m_1) u(n-m_2) \cdots u(n-m_{n+1}) |u(n-m_1)|^{k_1} |u(n-m_2)|^{k_2} \cdots |u(n-m_{n+1})|^{k_{n+1}} \\
&+ \cdots
\end{aligned} \tag{4.23}$$

It is clear from (4.23) that the new DPD architecture of MPM introduces many *high-order* cross terms and is much closer to the Volterra model than the classical MPM. As a result, the new DPD architecture is more powerful to model the inverse PA function and predistort the PA distortion.

On the other hand, please note that (4.23) is only the mathematical expansion by the Taylor Series in (4.17). The two MPMs of the proposed architecture still use the simple model structure as defined in (4.9), and their identification is also the same as the traditional MPM models. Hence it ensures that the new DPD scheme avoids the complexity of the Volterra model, and is as simple as the traditional MPM.

4.3.3 Coefficient Number

The number of coefficients of the Volterra model, the traditional MPM architecture, and the proposed MPM architecture are shown in Table 4-1. Since its number of coefficients increases very fast with nonlinear order and memory depth, the complexity of the Volterra model is very high. Meanwhile compared with the traditional MPM architecture, its number of coefficients is doubled for the proposed MPM architecture. However, since the coefficients of the proposed MPM are real-valued, while those of the traditional MPM are complex-valued, the complexity of the proposed MPM is similar to that of the traditional MPM.

<i>Model Type</i>	<i>Volterra Model</i>	<i>Traditional MPM</i>	<i>Proposed MPM</i>
<i>Coefficient Number</i>	$\frac{(K+M)!}{K!M!}$	KM	$2KM$

Table 4-1 The number of coefficients for different models, where K denotes nonlinear order, M denotes memory depth

4.3.4 Linearized Power Amplifier Block Diagram

Fig. 4-1 Linearized PA block diagram

Fig. 4-1 illustrates the block diagram of the linearized power amplifier (PA). $x(n)$ is the signal after predistorting $u(n)$, and is applied as input into PA. $y(n)$ is the final output. G is the gain of PA. The DPD trainer finds the inverse relation of the PA, and then is copied directly as digital predistortion. This is an indirect learning architecture [58].

For traditional architecture using memory polynomial model, only one MPM is used as the DPD. In the new architecture, two MPMs are used to predistort the amplitude and phase of the input signal separately. Thus, the overall architecture becomes (The parts of DAC, ADC, up-converter, down-converter and coupler are ignored here):

Fig. 4-2 New DPD architecture using MPM

Here $A(\bullet)$ and $\Phi(\bullet)$ are the amplitude inverse and phase inverse functions respectively, and both of them are memory polynomial models.

4.4 Simulation Results

In this section, we will show, by computer simulation, the performance of the new architecture using memory polynomial model.

The overall structure is based on Fig. 4-2. The baseband input signal is OFDM signal with 52 subcarriers, IFFT size of 64, and a bandwidth of 16.6MHz. The amplifier in the simulation is the behavior model characterizing the AM/AM and AM/PM relations of a Hittite GaAs InGaP HBT PA. The identification of the memory polynomial model of the DPD is carried out using 288001 data samples.

Both the traditional architecture [34] and the new one are simulated and compared. The traditional architecture is in the form of (4.8), and the new one is in the form of (4.10), (4.11) and (4.12). All the models are of 5-order nonlinearity and 3-tap memory effects, consisting of only odd-order items. However, for the traditional architecture, the coefficients of the MPM are in the form of complex numbers, while those of the new architecture are real-valued.

Fig. 4-3 shows the power spectral density (PSD) of the signals at the output side of the PA. Trace (a) shows the spectrum of the PA output linearized by the traditional DPD architecture and memory polynomial model. Results show that its spectral re-growth for frequencies in adjacent channel is around -30dB relative to the in-band level. Trace (b) shows the spectrum of the PA output linearized by the new DPD architecture. The spectral re-growth at adjacent channel is reduced to nearly -50dB . Trace (c) shows the spectrum of the original input signal. Compared with the traditional one, linearization using the new architecture improves the spectral re-growth by nearly 20 dB .

Adjacent Channel Power Ratio (ACPR) is a metric to quantify the out-of-band regrowth. The left ACPR and right ACPR are defined in (3.23) and (3.24) respectively. The results of ACPR performance over many frames are presented in Fig. 4-4. Trace (a) and (c) measure the ACPR values of different frames of the PA's output signal linearized by traditional architecture, while (b) and (d) measure the one linearized by the proposed architecture. It shows that the traditional DPD architecture has ACPR performance of around -31dB , while the new architecture can improve the ACPR to around -46dB .

Fig. 4-3 Normalized PSD of the signals, using memory polynomial model predistorter. (a) Predistorted by the traditional architecture. (b) Predistorted by the new architecture. (c) Original input signal

Fig. 4-4 ACPR over many frames. (a) Left ACPR with the traditional architecture. (b) Left ACPR with the new architecture. (c) Right ACPR with the traditional architecture. (d) Right ACPR with the new architecture.

The constellation and EVM of the output of the PA are also measured. Fig. 4-5(a) shows the constellation of the received signal from the traditional DPD architecture, which is obviously dispersed. On the contrary, in Fig. 4-5(b), the constellation has been greatly improved, with little dispersion and rotation. It suggests that the new DPD architecture can well compensate the AM/AM and AM/PM distortions caused by PA.

In Fig. 4-6, EVM is measured for different frames. The poorer EVM at the beginning is due to the initialization of the whole system. The new DPD architecture achieves about 15 dB improvement compared with the traditional one.

Fig. 4-5 constellation of the signals at the receiver after equalization. (a) Traditional architecture. (b) New architecture

4.5 Conclusion

This chapter presents a new DPD architecture for the memory polynomial predistortion model. It predistorts the amplitude and phase of the signal separately with two different memory polynomial models. The simulation results show that compared with the

traditional architecture, the new architecture can improve both ACPR and EVM of the PA output signals.

Besides the power amplifier, a new behavioral model will be proposed to characterize and predistort TX IQ imbalance of significant memory effects in the next chapter.

Fig. 4-6 EVM comparison over time at the receiver

Chapter 5 Frequency-Domain Predistortion of TX IQ Imbalance of Significant Memory Effects for OFDM Systems

5.1 Introduction

As modern wireless communication systems are evolving quickly with demand for higher data rates, signals are having wider and wider bandwidth. As a result, the IQ imbalance distortion shows more frequency selective effects, or longer memory effects. In this chapter, the long-memory TX IQ imbalance will be well characterized and predistorted by a new competent behavioral model.

Traditionally, TX/RX IQ imbalance is jointly compensated by some processing methods at the receiver, either in time-domain [94-98] or in frequency-domain [153, 154]. There are some efforts to jointly compensate TX/RX IQ imbalance with one other impairment [96, 105, 106]. However with the increased signal bandwidth, these post-compensation methods become increasingly complex.

Another approach is to isolate the impairments of the transmitter and the receiver and compensate them separately. Predistortion methods are used to compensate TX IQ imbalance. Some time-domain methods [107, 115-117, 120, 155] use digital filters placed before the DAC and modulator. The drawback of time-domain predistortion is that the

limitation on the length of the predistorter renders it unable to compensate long-memory IQ imbalance impairments. In this chapter, a new robust frequency-domain predistortion on long-memory TX IQ imbalance is proposed. Each subcarrier only needs one-tap complex-valued multiplier for itself and another one-tap complex-valued multiplier for its mirror frequency. Moreover, the problem is reduced to a matrix problem, and solved by robust numerical methods. It can efficiently and completely compensate long-memory TX IQ imbalance at various noise levels, without generating ISI distortion. Moreover, both its idea and its implementation are simple: It only involves multiplying the transmitted signals with some predistortion coefficients in frequency-domain. For the modern widely used OFDM systems [4, 139], by utilizing their existing function modules, only one multiplication module needs to be added. The numerical issue associated with the identification of the predistorter in the presence of noise and condition number problem [156, 157] is also analyzed.

The structure of the chapter is organized in the following way: In Section 5.2, the TX IQ imbalance model and *mirror frequency interference* are deducted and analyzed. In Section 5.3, the drawback of time-domain predistortion is demonstrated. In Section 5.4, the proposed frequency-domain TX IQ imbalance predistortion scheme is developed. Then in Section 5.5, the numerical issue is analyzed, the deficiency of Normal Equation is illustrated, and robust numerical methods are proposed to identify the predistorter, where the complexity analysis is also given. The simulation and real experiment results are shown in Section 5.6 and Section 5.7, respectively. And they are followed by the conclusion in Section 5.8.

Notations: The baseband or baseband equivalent time-domain signals are denoted by normal lowercase letters $s(n)$, while their frequency-domain counterparts are denoted by normal uppercase letters $S(k)$. For a given frequency bin $S(k)$, its mirror frequency bin is defined as $S(N-1-k)$ or $S(-k)$. Lowercase boldfaced letters denote time-domain column vectors $\mathbf{s}$, while the ones with the superscript f denote frequency-domain column vectors $\mathbf{s}^f$. Matrices are denoted by uppercase boldfaced letters. Time index within one OFDM frame is denoted by n , OFDM frame index is denoted by i , and frequency bin index is denoted by k . Hence $\mathbf{s}(i) = [s(0) \ s(1) \ \cdots \ s(N-1)]^T$ denotes the i th OFDM frame signal in time-domain, while $\mathbf{s}^f(i) = [S(0) \ S(1) \ \cdots \ S(N-1)]^T$ denotes its frequency-domain counterpart. And the vector $\mathbf{b}^f(k) = [S_1(k) \ S_2(k) \ \cdots \ S_L(k)]^T$ contains the values of the k th frequency bin over different L OFDM frames. The indices are also true for matrices. Note that the frame index i is sometimes neglected. The length of the training sequence is P OFDM frames, each having J subcarriers and N IFFT size. And the memory depth of the time-domain predistorter is M taps.

For operators, $(\cdot)^*$ and $\otimes$ denote the complex conjugate, and convolution, while $\hat{S}(k) = S^*(N-1-k) = S^*(-k)$ denotes the complex conjugate of the mirror frequency bin. $\mathbf{A}^*$, $\mathbf{A}^T$, $\mathbf{A}^H$, $\|\mathbf{A}\|_2$ denote the conjugate, transpose, conjugate-transpose, and 2-norm value respectively. If $\mathbf{A}$ is full-rank square matrix, $\mathbf{A}^{-1}$ denotes the matrix inverse operation. And $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$ denotes singular value decomposition. $\text{diag}\{\mathbf{x}\}$ denotes the

diagonal matrix with $\mathbf{x}$ being the diagonal. $\mathbf{0}_m$ is the null vector of m length with zeros in all entries. And $\mathbf{O}_{m \times n}$ is the $m \times n$ null matrix with zeros in all entries.

For more information of notations, please refer to **List of Principle Symbols** at the beginning of the thesis.

5.2 Signal and System Model

5.2.1 IQ Imbalance Model in Time Domain

Fig. 5-1 Block Diagram of TX IQ Imbalance: $\tilde{h}_I(t)$ and $\tilde{h}_Q(t)$ are the real-valued filters modeling the IQ imbalance of memory effects due to the imperfection of the DACs and the LPFs, and $a_I, \theta_I, a_Q, \theta_Q$ are the gain and phase imbalance owing to LO imbalance

The block diagram of TX IQ Imbalance is illustrated in Fig. 5-1. At the transmitter, the digital signal to be transmitted before digital to analog converter is $s(n) = s_I(n) + js_Q(n)$,

with T the symbol period. Let $\tilde{h}_s(t)$ denote the ideal shaping filter response, then the ideal LPF output is

$$\begin{aligned} s(t) &= s_I(t) + js_Q(t) \\ &= \sum_{i=-\infty}^{+\infty} s_I(n) \tilde{h}_s(t-nT) + j \sum_{i=-\infty}^{+\infty} s_Q(n) \tilde{h}_s(t-nT) \end{aligned} \quad (5.1)$$

However, due to the imperfection of the DACs and the LPFs along the I and Q paths, there is an imbalance of memory effects between the I and Q paths, which can be modeled by two filters after the ideal signal, i.e.,

$$\begin{aligned} x(t) &= x_I(t) + jx_Q(t) \\ &= s_I(t) \otimes \tilde{h}_I(t) + js_Q(t) \otimes \tilde{h}_Q(t) \end{aligned} \quad (5.2)$$

where $\tilde{h}_I(t)$ and $\tilde{h}_Q(t)$ are the real-valued filters modeling the IQ imbalance of memory effects.

Then going through the up-converters, the final transmitted signal gets another IQ imbalance distortion owing to the unequal gains and the deviation from 90° phase shift of the LOs. Hence, the transmitted signal is (neglecting the thermal noise of the transmitter)

$$y_p(t) = a_I x_I(t) \cos(2\pi f_c t + \theta_I) + a_Q x_Q(t) \sin(2\pi f_c t + \theta_Q) \quad (5.3)$$

where $a_I, \theta_I, a_Q, \theta_Q$ are the gain and phase imbalance owing to the imbalance of LOs.

Rearrange (5.3), the transmitted signal is

$$\begin{aligned} y_p(t) &= y_{p,I}(t) + y_{p,Q}(t) \\ &= (a_I x_I(t) \cos(\theta_I) + a_Q x_Q(t) \sin(\theta_Q)) \cos(2\pi f_c t) \\ &\quad + (a_Q x_Q(t) \cos(\theta_Q) - a_I x_I(t) \sin(\theta_I)) \sin(2\pi f_c t) \end{aligned} \quad (5.4)$$

And the equivalent low-pass transmitted signal is

$$\begin{aligned}
y(t) &= y_I(t) + jy_Q(t) \\
&= \left(a_I x_I(t) \cos(\theta_I) + a_Q x_Q(t) \sin(\theta_Q) \right) \\
&\quad + j \left(a_Q x_Q(t) \cos(\theta_Q) - a_I x_I(t) \sin(\theta_I) \right) \\
&= \left(s_I(t) \otimes a_I \tilde{h}_I(t) \cos(\theta_I) + s_Q(t) \otimes a_Q \tilde{h}_Q(t) \sin(\theta_Q) \right) \\
&\quad + j \left(s_Q(t) \otimes a_Q \tilde{h}_Q(t) \cos(\theta_Q) - s_I(t) \otimes a_I \tilde{h}_I(t) \sin(\theta_I) \right)
\end{aligned} \tag{5.5}$$

where it is obvious that due to the LO phase imbalance, there is crosstalk between I path and Q path. In addition, the imperfect isolation of the RF hardware devices also introduces another source of crosstalk. Hence all together, the equivalent low-pass transmitted signal after IQ imbalance distortion can be modeled by four real-valued filters:

$$\begin{aligned}
y(t) &= y_I(t) + jy_Q(t) \\
&= \left(s_I(t) \otimes h_{II}(t) + s_Q(t) \otimes h_{IQ}(t) \right) \\
&\quad + j \left(s_I(t) \otimes h_{QI}(t) + s_Q(t) \otimes h_{QQ}(t) \right)
\end{aligned} \tag{5.6}$$

Or its discrete form (from (5.1)):

$$\begin{aligned}
y(n) &= y_I(n) + jy_Q(n) \\
&= \left(s_I(n) \otimes h_{II}(n) + s_Q(n) \otimes h_{IQ}(n) \right) \\
&\quad + j \left(s_I(n) \otimes h_{QI}(n) + s_Q(n) \otimes h_{QQ}(n) \right)
\end{aligned} \tag{5.7}$$

Since $s_I(n) = (s(n) + s^*(n))/2$ and $s_Q(n) = (s(n) - s^*(n))/2j$, it can also be written as:

$$\begin{aligned}
y(n) &= \frac{1}{2} \left(h_{II}(n) + jh_{QI}(n) - jh_{IQ}(n) + h_{QQ}(n) \right) \otimes s(n) \\
&\quad + \frac{1}{2} \left(h_{II}(n) + jh_{QI}(n) + jh_{IQ}(n) - h_{QQ}(n) \right) \otimes s^*(n) \\
&= h_d(n) \otimes s(n) + h_m(n) \otimes s^*(n)
\end{aligned} \tag{5.8}$$

where $h_d(n)$ and $h_m(n)$ are the complex-valued direct filter and mirror filter, respectively.

5.2.2 OFDM Signal Distorted by TX IQ Imbalance

In a transmitter of an OFDM system, the data is firstly coded and mapped to some specific set of constellations in the frequency domain, and then converted to time domain by Inverse Fast Fourier Transform (IFFT). At the receiver side, the signal should also be converted to frequency domain by Fast Fourier Transform (FFT), before the data can be demapped from the constellations and decoded. Hence in general, by its nature, OFDM system provides a convenient platform to do frequency domain signal processing.

Let the frequency domain TX signal after coding and mapping denoted as:

$$\mathbf{s}^f = [S(0) \ S(1) \ \cdots \ S(N-1)]^T \quad (5.9)$$

where N is the FFT size. Throughout the chapter, a normal uppercase letter denotes a frequency domain variable of a specific frequency bin, while the superscript f denotes a frequency domain vector. Then the time domain TX signal is:

$$\begin{aligned} \mathbf{s} &= [\mathbf{s}_{N_{cp}}^{(cp)} \ \bar{\mathbf{s}}_N]^T \\ &= [s(N-N_{cp}) \ \cdots \ s(N-1) \ s(0) \ s(1) \ \cdots \ s(N-1)]^T \end{aligned} \quad (5.10)$$

where N_{cp} is the length of cyclic prefix (CP). And the signal without CP $\bar{\mathbf{s}}_N$ is

$$s(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S(k) e^{j \frac{2\pi nk}{N}}, \quad 0 \leq n \leq N-1 \quad (5.11)$$

Or in vector form:

$$\begin{cases} \bar{\mathbf{s}} = \mathbf{F}^* \mathbf{s}^f \\ \mathbf{s} = \mathbf{T}^{(cp)} \bar{\mathbf{s}} \\ \bar{\mathbf{s}} = \mathbf{C}^{(cp)} \mathbf{s} \end{cases} \quad (5.12)$$

where $\mathbf{F}$ is the unitary FFT matrix of size N :

$$[\mathbf{F}]_{nk} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi nk}{N}} \quad (5.13)$$

And the CP-adding matrix:

$$\mathbf{T}^{(cp)} = \begin{bmatrix} \mathbf{I}_{N_{cp} \times N} \\ \mathbf{I}_N \end{bmatrix} \quad (5.14)$$

where $\mathbf{I}_N$ is the identity matrix, while $\mathbf{I}_{N_{cp} \times N}$ is the last N_{cp} rows of $\mathbf{I}_N$. And the

CP-removing matrix is:

$$\mathbf{C}^{(cp)} = \begin{bmatrix} \mathbf{O}_{N \times N_{cp}} & \mathbf{I}_N \end{bmatrix} \quad (5.15)$$

Let the transmitted frequency-domain OFDM symbols be denoted as $\{\mathbf{s}^f(i)\}$, its corresponding time-domain symbols be denoted as $\{\mathbf{s}(i)\}$, and the final RF front-end low-pass equivalent OFDM symbols be denoted as $\{\mathbf{y}(i)\}$, where i is the index of the OFDM symbol. Let the vector form of the imbalance filters $h_d(n)$ be denoted as $\mathbf{h}_d = [h_d(0) \ \cdots \ h_d(L_d-1)]^T$, and the vector form of $h_m(n)$ is denoted as $\mathbf{h}_m = [h_m(0) \ \cdots \ h_m(L_m-1)]^T$, where L_d and L_m are the length of $h_d(n)$ and $h_m(n)$, respectively.

For convenience, we assume that the length of the OFDM symbol duration is longer than the maximum delay spread [139, 140, 158], i.e., $L_d < N + N_{cp}$ and $L_m < N + N_{cp}$.

And then from (5.8), it is obvious that the i th output OFDM symbol $\mathbf{y}(i)$ is solely determined by the $(i-1)$ th and i th input OFDM symbol $\mathbf{s}(i-1)$ and $\mathbf{s}(i)$. Specifically, from (5.8) and (5.12), the matrix form of the equivalent low-pass TX signal $\mathbf{y}(i)$ is:

$$\begin{aligned} \mathbf{y}(i) = & \tilde{\mathbf{H}}_d^c \mathbf{T}^{(cp)} \mathbf{F}^* \mathbf{s}^f(i) + \tilde{\mathbf{H}}_m^c \mathbf{T}^{(cp)} \mathbf{F} \mathbf{s}^{*f}(i) \\ & + \tilde{\mathbf{H}}_d^p \mathbf{T}^{(cp)} \mathbf{F}^* \mathbf{s}^f(i-1) + \tilde{\mathbf{H}}_m^p \mathbf{T}^{(cp)} \mathbf{F} \mathbf{s}^{*f}(i-1) \end{aligned} \quad (5.16)$$

where $\tilde{\mathbf{H}}_d^c$, $\tilde{\mathbf{H}}_m^c$, $\tilde{\mathbf{H}}_d^p$, and $\tilde{\mathbf{H}}_m^p$ are $(N+N_{cp}) \times (N+N_{cp})$ toeplitz matrices. A toeplitz matrix is a $N \times N$ square matrix that can be uniquely defined by its first row vector and its first column vector (For the definition of toeplitz matrix, please refer to Appendix C).

Here the first row and the first column of $\tilde{\mathbf{H}}_d^c$ are $\begin{bmatrix} h_d(0) & \mathbf{0}_{N+N_{cp}-1}^T \end{bmatrix}$ and

$\begin{bmatrix} \mathbf{h}_d^T & \mathbf{0}_{N+N_{cp}-L_d}^T \end{bmatrix}^T$, the first row and the first column of $\tilde{\mathbf{H}}_m^c$ are $\begin{bmatrix} h_m(0) & \mathbf{0}_{N+N_{cp}-1}^T \end{bmatrix}$ and

$\begin{bmatrix} \mathbf{h}_m^T & \mathbf{0}_{N+N_{cp}-L_m}^T \end{bmatrix}^T$, the first row and the first column of $\tilde{\mathbf{H}}_d^p$ are $\begin{bmatrix} \mathbf{0}_{N+N_{cp}-L_{TX,d}+1}^T & \hat{\mathbf{h}}_{TX,d}^T \end{bmatrix}^T$

and $\mathbf{0}_{N+N_{cp}}$, the first row and the first column of $\tilde{\mathbf{H}}_m^p$ are $\begin{bmatrix} \mathbf{0}_{N+N_{cp}-L_m+1}^T & \hat{\mathbf{h}}_m^T \end{bmatrix}^T$ and $\mathbf{0}_{N+N_{cp}}$,

where $\hat{\mathbf{h}}_d = [h_d(1) \ \cdots \ h_d(L_d-1)]^T$ and $\hat{\mathbf{h}}_m = [h_m(1) \ \cdots \ h_m(L_m-1)]^T$. Hence,

$$\tilde{\mathbf{H}}_d^c = \begin{bmatrix} h_d(0) & 0 & 0 & \cdots & 0 \\ h_d(1) & h_d(0) & 0 & \cdots & 0 \\ h_d(2) & h_d(1) & h_d(0) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_d(L_d-1) & h_d(L_d-2) & h_d(L_d-3) & \cdots & 0 \\ 0 & h_d(L_d-1) & h_d(L_d-2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & h_d(0) \end{bmatrix} \quad (5.17)$$

And note that $\tilde{\mathbf{H}}_m^c$ has the same form as $\tilde{\mathbf{H}}_d^c$.

$$\tilde{\mathbf{H}}_d^p = \begin{bmatrix} 0 & \cdots & 0 & h_d(1) & h_d(2) & \cdots & h_d(L_d-1) \\ 0 & \cdots & 0 & 0 & h_d(1) & \cdots & h_d(L_d-2) \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & h_d(1) \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \end{bmatrix} \quad (5.18)$$

And note that $\tilde{\mathbf{H}}_m^p$ has the same form as $\tilde{\mathbf{H}}_d^p$.

The second row in (5.16) is the ISI contribution the from previous OFDM symbol.

Note that the output $\mathbf{y}(i) = [\mathbf{y}_{N_{cp}}^{(cp)}(i) \quad \bar{\mathbf{y}}_N(i)]^T$ contains the cyclic prefix $\mathbf{y}_{N_{cp}}^{(cp)}(i)$, which can be removed by left-multiplying $\mathbf{C}^{(cp)}$ defined in (5.15):

$$\begin{aligned} \bar{\mathbf{y}}(i) &= \mathbf{C}^{(cp)} \mathbf{y}(i) \\ &= \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_d^c \mathbf{T}^{(cp)} \mathbf{F}^* \mathbf{s}^f(i) + \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_m^c \mathbf{T}^{(cp)} \mathbf{F} \mathbf{s}^{*f}(i) \\ &\quad + \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_d^p \mathbf{T}^{(cp)} \mathbf{F}^* \mathbf{s}^f(i-1) + \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_m^p \mathbf{T}^{(cp)} \mathbf{F} \mathbf{s}^{*f}(i-1) \end{aligned} \quad (5.19)$$

Here we further assume that the length of cyclic prefix is longer than the length of IQ imbalances [9, 140, 158], i.e., $L_d \leq N_{cp}$ and $L_m \leq N_{cp}$, then it is easy to show that

$$\begin{cases} \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_d^p = \mathbf{O}_{N \times N_{cp}} \\ \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_m^p = \mathbf{O}_{N \times N_{cp}} \end{cases} \quad (5.20)$$

Hence the last two items on the right-hand side of (5.19) are zeros. And after removing the CPs, $\bar{\mathbf{y}}(i)$ are free from ISI:

$$\begin{aligned}
\bar{\mathbf{y}}(i) &= \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_d^c \mathbf{T}^{(cp)} \mathbf{F}^* \mathbf{s}^f(i) + \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_m^c \mathbf{T}^{(cp)} \mathbf{F} \mathbf{s}^{*f}(i) \\
&= \mathbf{H}_d \mathbf{F}^* \mathbf{s}^f(i) + \mathbf{H}_m \mathbf{F} \mathbf{s}^{*f}(i)
\end{aligned} \tag{5.21}$$

where

$$\begin{cases} \mathbf{H}_d = \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_d^c \mathbf{T}^{(cp)} \\ \mathbf{H}_m = \mathbf{C}^{(cp)} \tilde{\mathbf{H}}_m^c \mathbf{T}^{(cp)} \end{cases} \tag{5.22}$$

A circulant matrix is a special toeplitz matrix that each row is only circle-shifting one element of its predecessor to the right. It can be uniquely determined by its first column (for details of circulant matrix, please refer to Appendix C). Here $\mathbf{C}^{(cp)}$ is given by (5.15), $\mathbf{T}^{(cp)}$ is given by (5.14), and $\tilde{\mathbf{H}}_d^c$ is given by (5.17). Hence from (5.14), (5.15), (5.17) and (5.22), it is easy to show that $\mathbf{H}_d$ is circulant matrix, whose first column vector is

$$\begin{bmatrix} \mathbf{h}_d^T & \mathbf{0}_{N-L_d}^T \end{bmatrix}^T : \tag{5.23}$$

$$\mathbf{H}_d = \begin{bmatrix} h_d(0) & 0 & \cdots & 0 & h_d(L_d-1) & \cdots & h_d(1) \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ h_d(L_d-2) & \cdots & h_d(0) & 0 & \cdots & 0 & h_d(L_d-1) \\ h_d(L_d-1) & \cdots & h_d(1) & h_d(0) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & h_d(L_d-1) & h_d(L_d-2) & \ddots & h_d(0) & 0 \\ 0 & \cdots & 0 & h_d(L_d-1) & \cdots & h_d(1) & h_d(0) \end{bmatrix}$$

Following the same deduction, $\mathbf{H}_m$ is also a circulant matrix, and its first columns is

$$\begin{bmatrix} \mathbf{h}_m^T & \mathbf{0}_{N-L_m}^T \end{bmatrix}^T .$$

5.2.3 IQ Imbalance in Frequency Domain

The frequency domain TX IQ imbalance can be obtained by doing FFT on both sides of (5.21), i.e.,

$$\begin{aligned}\mathbf{y}^f(i) &= \mathbf{F}\bar{\mathbf{y}}(i) \\ &= \mathbf{F}\mathbf{H}_d\mathbf{F}^*\mathbf{s}^f(i) + \mathbf{F}\mathbf{H}_m\mathbf{F}\mathbf{s}^{*f}(i)\end{aligned}\quad (5.24)$$

where the frequency response of the final RF front-end equivalent low-pass signal $\bar{\mathbf{y}}(i)$ of the i^{th} OFDM symbol is

$$\mathbf{y}^f(i) = \begin{bmatrix} Y_i(0) & Y_i(1) & \cdots & Y_i(N-1) \end{bmatrix}^T \quad (5.25)$$

The corresponding frequency response of the digital signal of the i^{th} OFDM symbol before DAC is

$$\mathbf{s}^f(i) = \begin{bmatrix} S_i(0) & S_i(1) & \cdots & S_i(N-1) \end{bmatrix}^T \quad (5.26)$$

Then define the frequency response of the mirror signal $\mathbf{s}^*(i)$ is

$$\hat{\mathbf{s}}^f(i) = \begin{bmatrix} \hat{S}_i(0) & \hat{S}_i(1) & \cdots & \hat{S}_i(N-1) \end{bmatrix}^T \quad (5.27)$$

According to its definition,

$$\hat{\mathbf{s}}^f(i) = \mathbf{F}\mathbf{s}^*(i) = \mathbf{s}^{*f}(i) \quad (5.28)$$

From (5.28), (5.24) can be written as:

$$\begin{aligned}\mathbf{y}^f(i) &= \mathbf{F}\bar{\mathbf{y}}(i) \\ &= \mathbf{\Lambda}_d\mathbf{s}^f(i) + \mathbf{\Lambda}_m\hat{\mathbf{s}}^f(i)\end{aligned}\quad (5.29)$$

where since $\mathbf{H}_d$ and $\mathbf{H}_m$ are circulant matrices, $\mathbf{\Lambda}_d$ and $\mathbf{\Lambda}_m$ are diagonal matrices.

Furthermore,

$$\begin{aligned}\mathbf{\Lambda}_d &= \text{diag} \left\{ \mathbf{F} \begin{bmatrix} \mathbf{h}_d^T & \mathbf{0}_{N-L_d}^T \end{bmatrix}^T \right\} \\ &= \text{diag} \left\{ \mathbf{h}_d^f \right\}\end{aligned}\tag{5.30}$$

And

$$\begin{aligned}\mathbf{\Lambda}_m &= \text{diag} \left\{ \mathbf{F} \begin{bmatrix} \mathbf{h}_m^T & \mathbf{0}_{N-L_m}^T \end{bmatrix}^T \right\} \\ &= \text{diag} \left\{ \mathbf{h}_m^f \right\}\end{aligned}\tag{5.31}$$

where $\mathbf{h}_d^f$ and $\mathbf{h}_m^f$ are the frequency response of the TX IQ imbalance filters $\mathbf{h}_d$ and $\mathbf{h}_m$.

5.3 Drawback of Time-Domain Predistortion

Fig. 5-2 shows the classical time-domain frequency-selective TX IQ imbalance predistortion scheme [115-117]. Fig. 5-3 shows the structure of the predistorter, which is a linear memory system having the similar structure as TX IQ imbalance. Here $c_{II}(n)$, $c_{QI}(n)$, $c_{IQ}(n)$, and $c_{QQ}(n)$ are all finite impulse response (FIR) filters

Considering $s_I(n) = (s(n) + s^*(n))/2$ and $s_Q(n) = (s(n) - s^*(n))/2j$, the output of the predistorter is:

$$\begin{aligned}x(n) &= x_I(n) + jx_Q(n) \\ &= \frac{1}{2} \left(c_{II}(n) + jc_{QI}(n) - jc_{IQ}(n) + c_{QQ}(n) \right) \otimes s(n) \\ &\quad + \frac{1}{2} \left(c_{II}(n) + jc_{QI}(n) + jc_{IQ}(n) - c_{QQ}(n) \right) \otimes s^*(n)\end{aligned}\tag{5.32}$$

Define

$$\begin{cases} c_d(n) = \frac{1}{2}(c_{II}(n) + jc_{QI}(n) - jc_{IQ}(n) + c_{QQ}(n)) \\ c_m(n) = \frac{1}{2}(c_{II}(n) + jc_{QI}(n) + jc_{IQ}(n) - c_{QQ}(n)) \end{cases} \quad (5.33)$$

Then the time-domain predistorter is in the following way:

$$x(n) = c_d(n) \otimes s(n) + c_m(n) \otimes s^*(n) \quad (5.34)$$

From (5.34) and (5.8), the final output signal is:

$$\begin{aligned} y(n) &= h_d(n) \otimes x(n) + h_m(n) \otimes x^*(n) \\ &= (h_d(n) \otimes c_d(n) + h_m(n) \otimes c_m^*(n)) \otimes s(n) \\ &\quad + (h_m(n) \otimes c_d^*(n) + h_d(n) \otimes c_m(n)) \otimes s^*(n) \end{aligned} \quad (5.35)$$

To eliminate the mirror frequency components, the second item in (5.35) should be zero, i.e.,

$$h_m(n) \otimes c_d^*(n) + h_d(n) \otimes c_m(n) = 0 \quad (5.36)$$

And the final output is free of TX IQ imbalance:

$$y(n) = (h_d(n) \otimes c_d(n) + h_m(n) \otimes c_m^*(n)) \otimes s(n) \quad (5.37)$$

Fig. 5-2 Time-domain TX IQ imbalance predistortion, where “Prd” stands for “predistortion”.

Fig. 5-3 Time-domain predistorter

We assume that the channel is changing slowly [139, 159-161], e.g., as a result of a low Doppler frequency [139]. Then the channel impulse response can be seen as time invariant in one OFDM frame duration [139, 158, 159], and modeled by a FIR filter $h_{ch}(n)$ [14, 140, 160]. Moreover, the impulse response of the receiver is also modeled by a FIR filter $h_{rx}(n)$ [158, 160]. Then the final received baseband signal is

$$y(n) = (h_d(n) \otimes c_d(n) + h_m(n) \otimes c_m^*(n)) \otimes h_{ch}(n) \otimes h_{rx}(n) \otimes s(n) \quad (5.38)$$

Let L_d^c , L_m^c , L_d , L_m , L_{ch} , L_{rx} denote the length of $c_d(n)$, $c_m(n)$, $h_d(n)$, $h_m(n)$, $h_{ch}(n)$, $h_{rx}(n)$ respectively and L_{cp} denotes the length of the cyclic prefix of OFDM frame. To avoid ISI impairment, the following equations should be satisfied:

$$\begin{cases} L_d^c + L_d + L_{ch} + L_{rx} \leq L_{cp} \\ L_m^c + L_m + L_{ch} + L_{rx} \leq L_{cp} \end{cases} \quad (5.39)$$

Or

$$\begin{cases} L_d^c \leq L_{cp} - L_d - L_{ch} - L_{rx} \\ L_m^c \leq L_{cp} - L_m - L_{ch} - L_{rx} \end{cases} \quad (5.40)$$

However, (5.40) sets the upper boundary of the length of the time-domain TX IQ imbalance predistorter. If either L_d^c or L_m^c is greater than the boundary, ISI impairment will be generated.

On the other hand, for a linear memory system, if it has long memory effects, then its time-domain inverse FIR filters should contain a significant number of taps, i.e., the length of its time-domain inverse FIR filter should be long [14, 162-168]. Hence since long-memory TX IQ imbalance has long memory length L_d and L_m ; and to compensate it, the length of the predistorter L_d^c and L_m^c should also be long.

Hence, it is contradictory and either time-domain predistorter cannot compensate long-memory TX IQ imbalance, or it will generate ISI impairment. This is the inherent disadvantage of time-domain predistorter.

The problem can be solved by using frequency-domain predistortion method. From (5.29) and (5.41), for the TX IQ imbalance, either of short or long memory, its effects in the frequency-domain are that the k^{th} output subcarrier is the weighted linear combination of the k^{th} input subcarrier and the complex conjugate of its mirror frequency, each having a complex-valued weight. Hence to predistort IQ imbalance in frequency domain, each subcarrier only needs a combination of one-tap multiplier for itself and another one-tap multiplier for its mirror frequency. In the next section, a robust frequency-domain predistorter is proposed.

5.4 Frequency-Domain Predistortion on TX IQ Imbalance of Long Memory Effects

For OFDM system, since there is already an IFFT operation in the transmitter, using frequency-domain TX IQ imbalance predistortion becomes very convenient. From (5.29), since Λ_d and Λ_m are diagonal matrices, the current frequency component $S(k)$ is only affected by its mirror frequency $S(-k)$. (5.29) can be decoupled for each individual frequency bin:

$$Y(k) = H_d(k) \cdot S(k) + H_m(k) \cdot \hat{S}(k) \quad (5.41)$$

where k is the frequency bin index, $H_d(k)$ and $H_m(k)$ are the frequency responses of the imbalanced filters $h_d(n)$ and $h_m(n)$, respectively. Hence in the frequency-domain, the effect of IQ imbalance impairments is that the k^{th} output subcarrier is the weighted linear combination of the k^{th} input subcarrier $S(k)$ and the complex conjugate of its mirror frequency $S(-k)$, whose weights are $H_d(k)$ and $H_m(k)$ respectively.

Equation (5.41) can be further written in matrix form:

$$\begin{bmatrix} Y(k) \\ \hat{Y}(k) \end{bmatrix}^T = \begin{bmatrix} S(k) \\ \hat{S}(k) \end{bmatrix}^T \cdot \mathbf{H}(k) \quad (5.42)$$

where the frequency-domain TX IQ imbalance matrix at frequency bin k is

$$\mathbf{H}(k) = \begin{bmatrix} H_d(k) & \hat{H}_m(k) \\ H_m(k) & \hat{H}_d(k) \end{bmatrix} \quad (5.43)$$

Fig. 5-4 Proposed frequency-domain long-memory TX IQ imbalance predistortion

The TX IQ imbalance predistorter is inserted between transmitted frequency-domain signal $S(k)$ and IFFT block as in Fig. 5-4. The TX IQ imbalance predistorter is linear. After predistortion, the signal becomes

$$\begin{bmatrix} X(k) \\ \hat{X}(k) \end{bmatrix}^T = \begin{bmatrix} S(k) \\ \hat{S}(k) \end{bmatrix}^T \cdot \mathbf{D}(k) \quad (5.44)$$

where the frequency-domain IQ imbalance predistortion matrix at frequency bin k is

$$\mathbf{D}(k) = \begin{bmatrix} d_{11}(k) & d_{12}(k) \\ d_{21}(k) & d_{22}(k) \end{bmatrix} \quad (5.45)$$

The scalar form of (5.44) is:

$$X(k) = d_{11}(k) \cdot S(k) + d_{21}(k) \cdot \hat{S}(k) \quad (5.46)$$

And after the IQ imbalance distortion, the frequency response of the final low-pass equivalent output signal is

$$\begin{bmatrix} Y(k) \\ \hat{Y}(k) \end{bmatrix}^T = \begin{bmatrix} S(k) \\ \hat{S}(k) \end{bmatrix}^T \mathbf{D}(k) \mathbf{H}(k) \quad (5.47)$$

From (5.47), in order to make TX signal free of IQ imbalance i.e., $Y(k) = S(k)$, the predistorter should be the inverse of the TX imbalance matrix, i.e.,

$$\mathbf{D}(k) = \mathbf{H}^{-1}(k) \quad (5.48)$$

Equation (5.48) is the frequency-domain TX imbalance predistorter. Furthermore, it is easy to show that

$$\begin{cases} d_{22}(k) = \hat{d}_{11}(k) \\ d_{12}(k) = \hat{d}_{12}(k) \end{cases} \quad (5.49)$$

Hence the TX predistortion matrix can be written as

$$\mathbf{D}(k) = \begin{bmatrix} D_d(k) & \hat{D}_m(k) \\ D_m(k) & \hat{D}_d(k) \end{bmatrix} \quad (5.50)$$

Furthermore, from (5.42) and (5.48),

$$\begin{bmatrix} Y(k) \\ \hat{Y}(k) \end{bmatrix}^T \cdot \mathbf{D}(k) = \begin{bmatrix} S(k) \\ \hat{S}(k) \end{bmatrix}^T \quad (5.51)$$

Hence $\mathbf{D}(k)$ can be obtained by directly solving (5.51).

5.5 Predistorter Identification and Sensitivity to Noise

The frequency-domain TX IQ imbalance predistorter $\mathbf{D}(k)$ can be identified by solving (5.51). However, the acquired $Y(k)$ suffers from noise perturbation. As a total number of P OFDM frames (whose FFT size is N) are gathered for each frequency bin, (5.51) can be converted to an over-determined equation:

$$\mathbf{A}(k) \cdot \mathbf{d}^f(k) \simeq \mathbf{b}^f(k) \quad (5.52)$$

where $\mathbf{A}(k)$ suffers from noise perturbation, and

$$\begin{cases} \mathbf{A}(k) = \begin{bmatrix} Y_1(k) & Y_2(k) & \cdots & Y_L(k) \\ \hat{Y}_1(k) & \hat{Y}_2(k) & \cdots & \hat{Y}_L(k) \end{bmatrix}^T \\ \mathbf{b}^f(k) = \begin{bmatrix} S_1(k) & S_2(k) & \cdots & S_L(k) \end{bmatrix}^T \end{cases} \quad (5.53)$$

And $\mathbf{d}^f(k)$ is the first column of $\mathbf{D}(k)$, i.e., $\mathbf{d}^f(k) = \begin{bmatrix} D_d(k) & D_m(k) \end{bmatrix}^T$

5.5.1 Accuracy of the Solution of Linear Equation

A general over-determined linear equation is written in the following form

$$\mathbf{Ax} = \mathbf{b} \quad (5.54)$$

Generally speaking, solving (5.54) is not as easy as it looks. Either the input or the output signal, or both of them, are affected by noise. And the noise can greatly degrade the accuracy of the calculated coefficients, or even lead to totally wrong values, depending on several factors. Michael T. Heath [157] pointed out that the final error in a solution is generally the sum of two types of errors – one is called propagated data error, and the other is called computational error. Propagated data error can be seen as the amount of noise propagated into the solution of the equation, which is determined by the identification task itself. It is a way to determine whether the identification problem itself is sensitive to the noise or not, and can be quantized by a factor called *condition number* [157]. Computational error can be seen as the error generated by the computing algorithm. It is a way to decide whether the computing algorithm is sensitive to noise or not. In general, a big error in the solution may be either due to a big propagated data error owing to the identification problem itself, or due to a big computation error owing to a sensitive algorithm.

In the case of linear model identification, e.g. linear TX IQ imbalance of long memory effects, the identification task itself is relatively insensitive to the input noise, or *well-conditioned*. Hence the propagated data error is small, and the accuracy of the solution of (5.52) is mainly determined by the computational error, or the robustness of the computing algorithm.

In the case of nonlinear system, e.g. deep nonlinear power amplifier, the *condition number* is often very large, and hence the identification task itself is often very sensitive to the input noise. The problem is *ill-conditioned*. Many publications reported that the main drawback of Volterra model is a large number of coefficients to identify, which is one of the reasons that attributes to the very large condition number – there are many items in a Volterra model, and these items are mutually correlated. In [169], P. C. Hansen analyzed rank-deficient and ill-conditioned problems. The identification of an *ill-conditioned* problem is usually very difficult, no matter what robust algorithms are applied. A better solution is to transform the problem into a *well-conditioned* problem, by changing another behavioral model.

In this chapter, since the subject is *well-conditioned* linear TX IQ imbalance, the main task is to reduce the computation error by applying a robust computing algorithm. The propagated data error and condition number problem will be discussed in the next chapter, where it becomes an *ill-conditioned* problem to characterize the unified system of PA and TX IQ imbalance by polynomial based models.

5.5.2 Norm Equation Algorithm and Its Deficiency

Equation (5.52) usually does not have a solution, unless $\mathbf{b}^f(k)$ lies in the space $\text{span}(\mathbf{A}(k))$. Least squares method is to find a L -vector $\mathbf{d}^f(k)$ that the difference

between $\mathbf{A}(k) \cdot \mathbf{d}^f(k)$ and $\mathbf{b}^f(k)$ are the least in the sense of sum of squared difference of every sample, or 2-norm of the error vector $\mathbf{e}(k) = \mathbf{b}^f(k) - \mathbf{A}(k)\mathbf{d}^f(k)$

$$\begin{aligned}\|\mathbf{e}\|_2 &= \sqrt{\mathbf{e}^H \mathbf{e}} \\ &= \sqrt{(\mathbf{b}^f - \mathbf{A}\mathbf{d}^f)^H (\mathbf{b}^f - \mathbf{A}\mathbf{d}^f)} \\ &= \sqrt{(\mathbf{b}^f)^H \mathbf{b}^f - (\mathbf{b}^f)^H \mathbf{A}\mathbf{d}^f - (\mathbf{d}^f)^H \mathbf{A}^H \mathbf{b}^f + (\mathbf{d}^f)^H \mathbf{A}^H \mathbf{A}\mathbf{d}^f}\end{aligned}\quad (5.55)$$

where the index is neglected for simplicity. To minimize the 2-norm of the error vector, the first-order derivative with respect to vector $\mathbf{x}$ should be equal to zero

$$\nabla \|\mathbf{e}\|_2^2 = -2\mathbf{A}^H \mathbf{b}^f + 2\mathbf{A}^H \mathbf{A}\mathbf{d}^f = 0 \quad (5.56)$$

Or

$$\mathbf{A}^H \mathbf{A}\mathbf{d}^f = \mathbf{A}^H \mathbf{b}^f \quad (5.57)$$

where (5.57) is often known as *normal equation*. So the solution of (5.57) in the sense of least 2-norm of error is

$$\mathbf{d}^f = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}^f \quad (5.58)$$

In the case of TX IQ imbalance modeling, due to the influence of the noise, the N -vector $\mathbf{b}^f$ in (5.52) is not in the L -dimension $\text{span}(\mathbf{A})$. So first project $\mathbf{b}^f$ onto $\text{span}(\mathbf{A})$ by an orthogonal projector $\mathbf{P} = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$, and then solve the equation in the closed space $\text{span}(\mathbf{A})$:

$$\begin{aligned}\mathbf{A}\mathbf{d}^f &= \mathbf{P}\mathbf{b}^f \\ &= \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}^f\end{aligned}\quad (5.59)$$

where we have the solution $\mathbf{d}^f = (\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}^f$. At the same time, the error vector can be obtained by projecting $\mathbf{b}^f$ onto the complementary space of $\text{span}(\mathbf{A})$:

$$\begin{aligned}
\mathbf{e} &= \mathbf{b} - \mathbf{A} \mathbf{d}^f \\
&= \mathbf{P}^\perp \mathbf{b}^f \\
&= (\mathbf{I} - \mathbf{P}) \mathbf{b}^f \\
&= (\mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H) \mathbf{b}^f
\end{aligned} \tag{5.60}$$

The solution of least squares is unique as long as matrix $\mathbf{A}$ is full rank. Otherwise, if $\text{rank}(\mathbf{A}) < L$, it is called *rank deficient*. In that case, there are unlimited number of solutions which are lying in the null space of $\text{span}(\mathbf{A})$, denoted by $\text{Null}(\mathbf{A})$. The common way to deal with rank deficient is to find the minimum 2-norm solution of least squares. In real applications of identification of TX IQ imbalance, owing to the noise, matrix $\mathbf{A}$ would always be full rank.

Using Normal Equation to solve an overdetermined equation is straightforward. However, the condition number of the Normal Equation is higher than that of the original matrix $\mathbf{A}$, i.e.,

$$\chi(\mathbf{A}^H \mathbf{A}) = [\chi(\mathbf{A})]^2 \tag{5.61}$$

where $\chi(\mathbf{A})$ denotes condition number of matrix $\mathbf{A}$.

With a high condition number, a small noise perturbation in $\mathbf{A}(k)$ can result in a large error in the solution $\mathbf{d}^f(k)$. Hence, the accuracy of the predistorter is quite sensitive to noise; and the noise-induced deviation in $\mathbf{A}(k)$ would generate large error in the coefficients of the predistorter and degrade the performance of predistortion.

The traditional way to solve the condition number problem in (5.52) is to increase the noise level in $\mathbf{A}(k)$ [160]. Although it can help in reducing the condition number value, the increased noise level itself can degrade the accuracy. As a result, the accuracy of the predistorter is still compromised.

5.5.3 Robust Identification Method

It is clear that in (5.52), only the left side $\mathbf{A}(k)$ suffers noise while the right side $\mathbf{b}^f(k)$ is free from noise. It is a data least squares (DLS) problem [170-175].

QR Method

Unitary transformation can be used to solve (5.52) [171, 172]. The property of a unitary transformation is that it maintains the 2-norm for the transformed vector. A unitary transformation matrix $\mathbf{Q}(k)$ can convert $\mathbf{b}^f(k)$ to a vector having all zero elements except the first one, i.e., $\mathbf{Q}^H(k)\mathbf{b}^f(k) = \begin{bmatrix} \|\mathbf{b}^f(k)\|_2 & 0 & \cdots & 0 \end{bmatrix}^T$, where $\mathbf{Q}(k)$ reserves the 2-norm of $\mathbf{b}^f(k)$. Multiply $\mathbf{Q}(k)$ on both sides of (5.52),

$$\begin{aligned} \mathbf{Q}^H(k)\mathbf{A}(k)\mathbf{d}^f(k) &\simeq \begin{bmatrix} \tilde{\mathbf{a}}_1^f(k) \\ \tilde{\mathbf{A}}_2(k) \end{bmatrix}^T \mathbf{d}^f(k) \\ &= \begin{bmatrix} \|\mathbf{b}^f(k)\|_2 \\ \mathbf{0} \end{bmatrix}^T \end{aligned} \quad (5.62)$$

where $\tilde{\mathbf{a}}_1^f(k)$ is the first row vector of $\mathbf{Q}^H(k)\mathbf{A}(k)$. Then (5.62) is equal to the following equations

$$\begin{cases} \tilde{\mathbf{a}}_1^f(k)\mathbf{d}^f(k) = \|\mathbf{b}^f(k)\|_2 \\ \tilde{\mathbf{A}}_2(k)\mathbf{d}^f(k) = \mathbf{0} \end{cases} \quad (5.63)$$

Firstly, solve the equation $\tilde{\mathbf{A}}_2(k)\mathbf{d}^f(k)=\mathbf{0}$, where $\mathbf{d}^f(k)$ lies in the Null space of $\tilde{\mathbf{A}}_2(k)$. Then $\mathbf{d}^f(k)$ is normalized to satisfy $\tilde{\mathbf{a}}_1^f(k)\mathbf{d}^f(k)=\|\mathbf{b}^f(k)\|_2$, so that the final solution is

$$\mathbf{d}^f(k)=\frac{\|\mathbf{b}^f(k)\|_2}{\tilde{\mathbf{a}}_1^f(k)\tilde{\mathbf{v}}_0^f(k)}\tilde{\mathbf{v}}_0^f(k) \quad (5.64)$$

where $\mathbf{v}_0^f(k)$ is a null vector of $\tilde{\mathbf{A}}_2(k)$.

QR-SVD Method

In real application, due to the noise in $\mathbf{A}(k)$, $\tilde{\mathbf{A}}_2(k)$ is not rank deficient, and does not have non-zero null vector $\mathbf{v}_0^f(k)$. In this case, a lower rank approximation can be generated by *singular value decomposition* (SVD) [176]. The approximation $\tilde{\mathbf{A}}'_2(k)$ has the minimum difference in the sense of 2-norm value, i.e. $\min\|\tilde{\mathbf{A}}'_2(k)-\tilde{\mathbf{A}}_2(k)\|_2$. Suppose $\tilde{\mathbf{A}}_2(k)$ is of full column rank L , and can be decomposed by

$$\begin{aligned}\tilde{\mathbf{A}}_2(k) &= \mathbf{U}(k)\Sigma(k)\mathbf{V}^H(k) \\ &= \sigma_1\mathbf{u}_1\mathbf{v}_1^H + \sigma_2\mathbf{u}_2\mathbf{v}_2^H + \cdots + \sigma_L\mathbf{u}_L\mathbf{v}_L^H \\ &= \sigma_1\mathbf{E}_1 + \sigma_2\mathbf{E}_2 + \cdots + \sigma_L\mathbf{E}_L\end{aligned} \quad (5.65)$$

Then its 1-rank lower approximation is to delete the item associated with the minimum singular value:

$$\begin{aligned}\tilde{\mathbf{A}}'_2(k) &= \mathbf{U}(k)\Sigma'(k)\mathbf{V}^H(k) \\ &= \sigma_1\mathbf{u}_1\mathbf{v}_1^H + \sigma_2\mathbf{u}_2\mathbf{v}_2^H + \cdots + \sigma_{L-1}\mathbf{u}_{L-1}\mathbf{v}_{L-1}^H \\ &= \sigma_1\mathbf{E}_1 + \sigma_2\mathbf{E}_2 + \cdots + \sigma_{L-1}\mathbf{E}_{L-1}\end{aligned} \quad (5.66)$$

where $\mathbf{E}_i = \mathbf{u}_i \mathbf{v}_i^H$ is $N \times L$ matrix and has rank 1. And now $\mathbf{v}_0^f(k)$ is a non-zero null vector of $\tilde{\mathbf{A}}'_2(k)$.

There is a simpler way to avoid doing the lower rank approximation: Find the right singular vector $\tilde{\mathbf{v}}_L^f(k)$ corresponding to the minimum singular value σ_L of $\tilde{\mathbf{A}}_2$, and the QR-SVD solution of (5.52) is

$$\mathbf{d}^f(k) = \frac{\|\mathbf{b}^f(k)\|_2}{\tilde{\mathbf{a}}_1^f(k) \tilde{\mathbf{v}}_L^f(k)} \tilde{\mathbf{v}}_L^f(k) \quad (5.67)$$

Note that if the null space of $\tilde{\mathbf{A}}_2(k)$ is orthogonal to vector $\tilde{\mathbf{a}}_1^f(k)$, (5.63) does not have a non-zero solution, and hence (5.52) does not have a solution.

5.5.4 Complexity Analysis

Here the complexity is valued in term of floating point operation (flop). Assume the training sequence has the length of P OFDM frames, each having J subcarriers and N IFFT size. Then for (5.52) and a specific frequency bin k , $\mathbf{A}(k)$ is a $P \times 2$ matrix, $\mathbf{d}^f(k)$ and $\mathbf{b}^f(k)$ are 2×1 and $P \times 1$ vectors, respectively. For the complexity of QR algorithm, firstly the QR decomposition of $\mathbf{b}^f(k)$ requires $8P^2 - 16P + 10$ flops by the classical Householder method. Then to compute the null vector $\tilde{\mathbf{A}}_2(k) \mathbf{d}^f(k) = \mathbf{0}$, another QR decomposition is carried out with $8P^2 - 32P + 34$ flops, and computing (5.64) requires 6 flops. Hence QR algorithm requires $16JP^2 - 48JP + 50J$ flops. For the complexity of QR-SVD algorithm, the SVD operation requires $16P + 48$ flops by the classical Golub-Reinsch SVD method [177], and computing (5.67) requires 6 flops. Hence QR-SVD algorithm requires $8JP^2 + 64J$ flops. As a comparison, the state-of-the-art

time-domain predistortion in [117] requires $O(P^3N^3)$ flops to identify its compensator. Since $N > J$, the complexity of the two proposed algorithms is significantly lower than the state-of-the-art time-domain solution. The complexity analysis is shown in Table 5-1.

Algorithm	Flops
Frequency-domain QR method	$16JP^2 - 48JP + 50J$
Frequency-domain QR-SVD method	$8JP^2 + 64J$
State-of-the-art time-domain method in [117]	$O(P^3N^3)$

Table 5-1 Identification Complexity Comparison

From (5.41), the proposed predistorter is a bank of complex-valued weight pairs $H_d(k)$ and $H_m(k)$. Hence For a signal of J subcarriers, the total number of weights is $2J$. On the other hand, for the time-domain predistorter in [117] with memory depth of M -taps, the total number of complex-valued coefficients is $2M$. The comparison is shown in Table 5-2. If $J < M$, the frequency-domain predistorter has less coefficients; otherwise, the time-domain predistorter has less coefficients. However, as discussed above and shown in Table 5-1, even when $J > M$, the identification complexity of the frequency-domain predistorter is still lower than that of the time-domain predistorter.

predistorter	Number of coefficients
Proposed frequency-domain predistorter	$2J$
time-domain predistorter in [117]	$2M$

Table 5-2 Comparison of the number of coefficients: J is the number of subcarriers of the OFDM signal, while M is the length of time-domain predistorter

5.6 Simulation Results

An OFDM system based on WiFi [2, 167] is simulated. The IFFT size of the OFDM symbol is 256. There are 52 meaningful subcarriers, the space between each two adjacent subcarriers is 0.3125MHz, and the total bandwidth is 16.6 MHz. The 64-QAM constellation schemes are simulated. The TX IQ imbalance distortion is modeled by the scheme in Fig. 5-5. All four frequency selective branches $h_{II}(n)$, $h_{IQ}(n)$, $h_{QI}(n)$, $h_{QQ}(n)$ are real-valued low-pass filters, all having long memory depth. The EVM defined in Section 3.5.2 is used as the criteria to evaluate the performance:

$$EVM = \sqrt{\frac{\sum_{l=1}^L \sum_{k=1}^N |Y_l(k) - S_l(k)|^2}{\sum_{l=1}^L \sum_{k=1}^N |Y_l(k)|^2}} \quad (5.68)$$

where $Y_l(k)$ and $S_l(k)$ are the signals as defined in (5.47), and L is the number of OFDM symbols, N is the IFFT size. EVM performance of the proposed frequency domain TX IQ imbalance predistortion scheme is compared with the state-of-the-art time

domain predistortion scheme in [117], the length of which is 10 taps. The performance of the system without IQ imbalance compensation and the ideal case of no IQ imbalance are also included for reference. Since $\tilde{\mathbf{A}}_2(k)$ is full rank due to noise, QR-SVD method is used to identify the pre-compensator. And the time alignment method elaborated in Section 3.2.6 is used to compensate the delay mismatch between the input and output signals in this section and the next section.

Fig. 5-5 Behavioral model of the TX IQ imbalance distortion in the simulation

$h_{II}(n)$	0.0223, 0.0357, 0.0122, -0.0169, -0.0568, -0.0454, 0.0233, 0.1468, 0.2588, 0.3079, 0.2588, 0.1468, 0.0233, -0.0454, -0.0568, -0.0169, 0.0122, 0.0357, 0.0223
$h_{IQ}(n)$	-0.0011, -0.0012, -0.0002, 0.0025, 0.0063, 0.0096, 0.0109, 0.0096, 0.0063, 0.0025, -0.0002, -0.0012, -0.0011

$h_{QI}(n)$	-0.0011, -0.0017, -0.0017, -0.0002, 0.0028, 0.0063, 0.0086, 0.0086, 0.0063, 0.0028, -0.0002, -0.0017, -0.0017, -0.0011
$h_{QO}(n)$	0.0217, 0.0361, 0.0121, -0.0168, -0.0566, -0.0457, 0.0234, 0.1465, 0.2591, 0.3078, 0.2591, 0.1465, 0.0234, -0.0457, -0.0566, -0.0168, 0.0121, 0.0361, 0.0217

Table 5-3 Coefficients of the IQ imbalance model in “EVM vs SNR” simulation

Firstly, the EVM performance versus signal-to-noise-ratio values is evaluated. The IQ imbalance model in Fig. 5-5 is fixed, and its coefficients are shown in Table 5-3. The results are shown in Fig. 5-6 for 64-QAM constellations. In the figure, “ideal w/o IQ imbalance” refers to a transmitter without IQ imbalance distortion, which works as a benchmark. And “w/o predistort” refers to a transmitter with IQ imbalance distortion, but there is no compensation scheme. And “time-domain predistort” refers to the transmitter with IQ imbalance distortion, and there is the time-domain TX IQ imbalance predistortion. And finally, “frequency-domain predistort” refers to the transmitter with IQ imbalance distortion, and there is the frequency-domain TX IQ imbalance predistortion proposed in this chapter.

The time-domain TX IQ imbalance predistortion has some limit in the EVM performance, where further increase in the SNR does not help to significantly reduce the EVM value. This is because the time-domain predistorter cannot completely compensate TX IQ imbalance – there is still some residual IQ imbalance effects after compensation. On the contrary, the EVM of the proposed frequency-domain predistortion keeps

decreasing as SNR is increased. It almost coincides with those of the “ideal w/o IQ imbalance” case.

Fig. 5-6 EVM performance versus SNR values, under 64 QAM scheme

Influence of memory length of the TX IQ imbalance distortion on the EVM performance is also evaluated and shown in Fig. 5-7 for 64-QAM constellations. The SNR is fixed at 45 dB. The memory length of the four frequency selective branches $h_I(n)$, $h_{IQ}(n)$, $h_{QI}(n)$, $h_Q(n)$ of the TX IQ imbalance model in Fig. 5-5 is varying from 0.125us to 0.8us (Consequently, their coefficients are also changed). Note that the time duration of the cyclic-prefix of typical OFDM symbol is 0.8us according to IEEE 802.11a. It is demonstrated that the time-domain predistorter cannot compensate long-memory TX IQ imbalance impairments – The EVM performance of time-domain predistortion degrades with increase in memory depth in IQ imbalance. On the contrary, the proposed

frequency-domain TX IQ imbalance predistortion shows quite constant EVM performance.

Fig. 5-7 EVM performance versus memory length of the TX IQ imbalance distortion under 64 QAM scheme

The effects of compensation on each subcarrier are also compared in Fig. 5-8, where the EVM performance of all the 52 subcarrier is calculated and compared. Firstly, since the IQ imbalance impairments are frequency-selective, without predistortion, the EVM values of different subcarriers are different in “w/o predistort”. Then from the EVM performance of “time-domain predistort”, it is clear that the time-domain predistortion method also shows frequency-selective effects. On the contrary, the EVM performance of “frequency-domain predistort” is almost equal over all the subcarriers, demonstrating that the proposed frequency-domain predistorter has consistent performance over the whole bandwidth.

Then, the condition number is also compared between the proposed frequency-domain predistorter and the time-domain predistorter, as illustrated in Fig. 5-9. The condition number is increasing very fast with the SNR. The high value of condition number makes the time-domain predistorter very sensitive to noise, resulting in inaccurate coefficients. Meanwhile, the condition number of the proposed frequency-domain predistorter is the maximum value among all the subcarriers. It is clear that the condition number of the frequency-domain predistorter is very small. Hence the frequency-domain predistorter can overcome the condition number problem, and obtain accurate coefficients.

Last but not least, the proposed method is compared with the other state-of-the-art frequency-domain TX IQ imbalance predistortion methods published in [118, 119]. The comparison results are shown in Fig. 5-10, where the EVM performance of different lengths of training sequences is compared. In the figure, “proposed method” refers to the method proposed in this chapter, while “zou2008” and “Mcpherson2011” refer to those in [118] and [119] respectively. The simulation is done under 32 dB SNR, and averaged over 300 runs. The “proposed method” shows superior performance in the sense of fast convergence – “proposed method” converges at around 20 OFDM symbols, while “zou2008” and “Mcpherson2011” converges at around 60 ~ 70 OFDM symbols. Hence “proposed method” converges 3 times faster than “zou2008” and “Mcpherson2011”. The reason is that the identification method proposed in Section 5.5.3 is more robust and less sensitive to noise, and as a result, it needs shorter training sequence.

Fig. 5-8 Comparison of the EVM performance over each subcarrier

Fig. 5-9 Condition number comparison between the proposed frequency-domain predistorter and the time-domain predistorter.

Fig. 5-10 Comparison with other frequency-domain TX IQ imbalance predistortion methods. Different lengths of training sequence are swept, where the length of training sequence is in the number of OFDM symbols.

5.7 Experiment Results

The performance of the proposed frequency-domain TX IQ imbalance predistortion is also evaluated in the real experiment. The setup is shown in Fig. 5-11. The signal generator (Agilent ESG E4438C) consists of the I-Path/Q-Path DACs, the reconstruction filters and the up-converters, as well as other analog devices. The RF carrier frequency is 2.4 GHz. In order to avoid extra IQ imbalance in the feedback path, the RF signal is down-converted to intermediate frequency of 100MHz, and received and digitized by the oscilloscope (Infiniium 54832D DSO). Then, the digital IF signal is down-converted to baseband by the digital demodulator. By this way, extra IQ imbalance distortion in the feedback path is avoided. After FFT operation, the received signal $Y(k)$ and the

transmitted signal $S(k)$ are used to identify the frequency domain TX IQ imbalance predistorter.

Fig. 5-11 Experiment setup for frequency domain TX IQ imbalance predistortion.

$h_{II}(n)$	0.0260, 0.0222, 0.0151, -0.0050, -0.0310, -0.0481, -0.0400, 0.0021, 0.0734, 0.1546, 0.2190, 0.2435, 0.2190, 0.1546, 0.0734, 0.0021, -0.0400, -0.0481, -0.0310, -0.0050, 0.0151, 0.0222, 0.0260
$h_{IQ}(n)$	-0.0010, -0.0009, -0.0006, 0.0006, 0.0026, 0.0050, 0.0072, 0.0085, 0.0085, 0.0072, 0.0050, 0.0026, 0.0006, -0.0006, -0.0009, -0.0010

$h_{QI}(n)$	-0.0009, -0.0012, -0.0013, -0.0007, 0.0008, 0.0028, 0.0051, 0.0068, 0.0074, 0.0068, 0.0051, 0.0028, 0.0008, -0.0007, -0.0013, -0.0012, -0.0009
$h_{QQ}(n)$	0.0262, 0.0225, 0.0156, -0.0045, -0.0305, -0.0476, -0.0397, 0.0024, 0.0737, 0.1550, 0.2195, 0.2440, 0.2195, 0.1550, 0.0737, 0.0024, -0.0397, -0.0476, -0.0305, -0.0045, 0.0156, 0.0225, 0.0262

Table 5-4 Coefficients of the additional IQ imbalance model in the experiment

To simulate a low cost vector modulator that has poor IQ imbalance, additional TX IQ imbalance distortion is deliberately added in the software before the signal generator. It is modeled by the same scheme as illustrated in Fig. 5-5, and its coefficients are illustrated in Table 5-4. The signal used in the experiment is 16.6 MHz OFDM signal based on WiFi [2, 167]. 10,000 samples are used in the identification of the predistorter. Another 10,000 samples are used to test the EVM performance. The performance is compared with that of the state-of-the-art time-domain predistortion scheme [117], and the case of no IQ imbalance predistortion. The length of the time-domain predistorter is 10 taps.

<i>schemes</i>	<i>EVM (64-QAM)</i>
no predistortion	-22.974 dB
time-domain predistortion	-34.989 dB
frequency-domain predistortion	-41.623 dB

Table 5-5 EVM performance of different schemes in experiment measurements

The specifications of the signal are: The IFFT size of the OFDM symbol is 256, the number of the meaningful subcarriers is 52, the space between each two adjacent subcarriers is 0.3125MHz, and the total bandwidth is 16.6 MHz. The experiment is done with SNR of around 48 dB. Since $\tilde{\mathbf{A}}_2(k)$ is full rank due to noise, QR-SVD method is used to identify the predistorter. The evaluation results are shown in Table 5-5. It is obvious that the proposed frequency-domain predistortion scheme outperforms the time-domain predistortion by about 7 dB for 64-QAM constellations. The 64-QAM constellation comparison obtained from the experiment measurements with and without the proposed frequency-domain predistortion scheme is shown in Fig. 5-12.

Fig. 5-12 Constellation comparison in real experiments: The left figure is the 64-QAM constellation suffering from TX IQ imbalance of long-memory; the right figure is the corresponding constellation after using the proposed frequency-domain predistortion to compensate.

5.8 Conclusion

In this chapter, a new robust frequency-domain TX predistorter is proposed to compensate TX IQ imbalance of long memory effects. It is to convert TX IQ imbalance

problem to a matrix problem, and solve it by robust methods. Unlike the time-domain methods, the proposed predistortion scheme can perfectly compensate the long memory effects in the IQ imbalance. And compared with other frequency-domain predistortion methods, it is robust and requires less training sequence. It also takes advantage of the OFDM system structure, where the IFFT and FFT operation already exist, and hence no additional Fourier transform operation is needed. The condition number problem in the Normal Equation can degrade the accuracy of the identification of the predistorter and result in poor predistorter performance. Based on DLS problem, two numerical methods are proposed to counter the condition number problem. And due to noise, only QR-SVD method is applicable in real applications.

Chapter 6 Novel Neural Network Model of Power Amplifier + IQ Imbalance

6.1 Introduction

In the earlier chapters, new behavioral models for PA and TX IQ imbalance distortions are proposed and used to predistort each distortion separately. In this chapter, efforts will be made to merge the two distortions into one black box, and characterize them by a single behavioral model. In this chapter, the TX IQ Imbalance and power amplifier distortions are treated as a single black box, and modeled in one behavioral model -- *Rectangular structured Focused Time-Delay Neural Network* (RSFTDNN). Here the advantages of RSFTDNN are: (1) It provides a unified view of the impairments of the transmitter. (2) It can accurately characterize a deeply nonlinear communication system. (3) The network structure is simple, and easy to be implemented, and it can save computation resources.

The proposed model is designed and specified according to the characteristics of power amplifier and TX IQ imbalance. The proposed neural network is double input (real part and imaginary part separately) and single output (complex-valued). In order to model the “bend-down” AM/AM characteristics of a power amplifier, the activation function should be saturated for large input. And to model the memory effects, the proposed neural

network uses both the current and the past inputs. Furthermore, *condition number* problem on polynomial based model is also discussed and analyzed.

The structure of the chapter is arranged as follows: Section 6.2 illustrates the scheme of uniformly modeling PA distortion and IQ imbalance. Section 6.3 analyzes the limitation of *condition number* problem on polynomial based models. Section 6.4 proposes and analyzes the RSFTDNN model. And Section 6.5 shows the experiment results. Finally, Section 6.6 gives the conclusion.

6.2 Unified Model

The overall diagram of a typical wireless transmitter is shown in Fig. 2-3. The distortion of TX IQ imbalance refers to the impairments or difference between $s(n)$ and $x_l(n)$ (digitized version of $x_l(t)$); and the distortion of the power amplifier refers to the impairments or difference between $x_l(n)$ and $y_l(n)$ (digitized version of $y_l(t)$). Traditionally, these two distortions are modeled and simulated separately by two independent models. One possible reason for treating these two distortions separately may be that they are caused by two different physical mechanisms. However, from the viewpoint of system-level simulation, all the underneath physical characteristics are ignored, while only the output signals are of concern. Hence it is natural and sensible to merge these two distortions into one unified black box, using one mathematical equation to model the relationship between its input $s(n)$ and output $y_l(n)$. This saves computation resources, and presents a unified view of the transmitter impairments.

6.3 Limitation on Polynomial Based Models

As discussed in Section 5.5.1, the error of model identification is the sum of computation error and propagated data error. While the former is related to the robustness of the computing algorithms, the latter is related to the identification problem itself – whether the identification problem itself is sensitive to noise or not. Propagated data error can be quantized by condition number. This section will discuss the propagated data error, and the condition number problem on polynomial based models.

Polynomial based models and neural network models are two general classes of models which can be used to characterize the behavior of a system. The class of polynomial based models can only model mildly nonlinear system [25, 27], because it is difficult to accurately identify the coefficients with very high nonlinear order. This is due to the fact that the condition number increases very quickly with increasing nonlinear order and memory depth. And a large condition number means inaccurate identification results as demonstrated in the derivation below.

One important character of polynomial based model is that the output of the model is in linear relationship with the coefficients. Hence usually the way to identify the coefficients is to solve the linear equation:

$$\mathbf{Ax} \simeq \mathbf{b} \quad (6.1)$$

where the matrix $\mathbf{A}$ contains all the linear and nonlinear orders of the input samples, vector $\mathbf{b}$ is the output samples, and the vector variable $\mathbf{x}$ is the coefficients to be identified. Usually, the number of samples is much larger than the number of coefficients, resulting in a least squares (LS) problem.

6.3.1 Norm and Condition Number

The *2-norm* of an L -dimensional vector $\mathbf{x}$ is defined

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^L |x_i|^2} \quad (6.2)$$

And accordingly the 2-norm of a matrix $\mathbf{A}$ is defined

$$\begin{aligned} \|\mathbf{A}\|_2 &= \sup_{\mathbf{x} \neq 0} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} \\ &= \sigma_{\max} \end{aligned} \quad (6.3)$$

And *Frobenius* norm of a matrix $\mathbf{A}_{N \times L}$ is defined

$$\begin{aligned} \|\mathbf{A}\|_F &= \sqrt{\sum_{i=1}^N \sum_{j=1}^L |a_{ij}|^2} \\ &= \sqrt{\text{tr}(\mathbf{A}^H \mathbf{A})} \\ &= \sqrt{\sum_{i=1}^{\min(N,L)} \sigma_i^2} \end{aligned} \quad (6.4)$$

where $\{\sigma_i\}_{\min(N,L)}$ are the singular values of $\mathbf{A}$.

Using 2-norm, the condition number of a square matrix $\mathbf{A}_{N \times N}$ is defined

$$\begin{aligned} \chi(\mathbf{A}) &= \|\mathbf{A}\|_2 \cdot \|\mathbf{A}^{-1}\|_2 \\ &= \frac{|\lambda|_{\max}}{|\lambda|_{\min}} \end{aligned} \quad (6.5)$$

where $\{\lambda_i\}_L$ is the eigenvalues of $\mathbf{A}_{N \times L}$. If $\mathbf{A}$ is not full rank, it is defined that

$$\chi(\mathbf{A}) = +\infty.$$

For rectangular matrix $\mathbf{A}_{N \times L}$, where $N > L$, the condition number is defined

$$\begin{aligned}\chi(\mathbf{A}) &= \|\mathbf{A}\|_2 \cdot \|\mathbf{A}^\dagger\|_2 \\ &= \frac{\sigma_{\max}}{\sigma_{\min}}\end{aligned}\tag{6.6}$$

where $\{\sigma_i\}_L$ is the singular values of $\mathbf{A}_{N \times L}$ and $\|\mathbf{A}\|_2$ is 2-norm of matrix $\mathbf{A}$. And if $\mathbf{A}$ is not full rank, it is set that $\chi(\mathbf{A}) = +\infty$.

Condition number is a way to measure the sensitivity of the model accuracy to noise. If the input signal is the baseband digital signal before DAC, and the output signal is the down converted signal of DUT output after ADC, then the noise only lies in vector $\mathbf{b}$.

The condition number $\chi(\mathbf{A})$ is totally determined by matrix $\mathbf{A}$. In case of the a deep nonlinear and significant memory system like a wide-band power amplifier plus TX IQ imbalance, the identification is usually an *ill-conditioned* task, and $\chi(\mathbf{A})$ increases fast with increasing nonlinear order or memory depth.

6.3.2 Limitation on the Accuracy of Polynomial Based Models

The angle between the N -dimensional vector $\mathbf{b}$ and the space $\text{span}(\mathbf{A})$ is defined

$$\begin{aligned}\cos \theta &= \frac{\|\mathbf{P}^\perp \mathbf{b}\|_2}{\|\mathbf{b}\|_2} \\ &= \frac{\|\mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H \mathbf{b}\|_2}{\|\mathbf{b}\|_2} \\ &= \frac{\|\mathbf{A} \mathbf{x}\|_2}{\|\mathbf{b}\|_2}\end{aligned}\tag{6.7}$$

where $\mathbf{x}$ is the least squares solution. The angular between $\mathbf{b}$ and $\text{span}(\mathbf{A})$ measures how large the deviation is generated. The deviation can be generated by noise, or by approximation error of the model.

Theorem 6-1 For full column rank $N \times L$ matrix $\mathbf{A}$, and if noise is only present in vector $\mathbf{b}$, the upper boundary of the sensitivity of the LS solution is calculated [157] as:

$$\frac{\|\Delta \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\chi(\mathbf{A})}{\cos \theta} \frac{\|\Delta \mathbf{b}\|_2}{\|\mathbf{b}\|_2} \quad (6.8)$$

Proof. The proof is also given in [49]. $\square$

Equation (6.8) prevails in all the polynomial models that have linear relationship between their outputs and their coefficients. It shows that the error in the coefficients $\mathbf{x}$ is highly correlated to the condition number $\chi(\mathbf{A})$, with its upper boundary being linearly proportional to the condition number. Since its condition number increases quickly with nonlinear order and memory depth, the error of the polynomial model also increases dramatically. As a result, polynomial based models can only characterize systems of mild nonlinearity and short memory effects [25, 27].

On the contrary, neural network is good at modeling even a highly nonlinear system with long memory effects [27, 70, 71]. With a proper number of layers and neurons, the neural network can characterize a system with arbitrary accuracy [70].

Furthermore, from Fig. 2-3, it is clear that to characterize TX IQ imbalance and PA distortions together, the model should have double inputs (real $s_I(n)$ and imaginary $s_Q(n)$ respectively). However, the classical polynomial model is accustomed to model single input single output (SISO) system, and hence some modification or extension are necessary to model such double input system, which may very well increase the complexity and computation cost. On the contrary, neural network has natural flexibility in

the number of inputs and outputs, and can model such double input systems quite straightforwardly.

Last but not least, neural network enjoys more freedom: i.e. it can accommodate varying number of layers or neurons in each layer, changes in the type of activation function, or even changes in the connections or structure. Hence, neural network may be a better choice to model the unified system consisting of TX IQ imbalance and highly nonlinear PA.

6.4 Proposed Neural Network Model

6.4.1 Prototype of Unified Model

The IQ imbalance and PA distortions in the transmitter are shown in Fig. 6-1. The behaviors of TX IQ imbalance and PA impairments can be further abstracted by two cascaded separate blocks as shown in Fig. 6-2. Now we will deduce the form of the unified behavioral model – The form of relationship between $s(n)$ and $y(n)$.

Firstly, the relationship between $s(n)$ and $x(n)$ is TX IQ imbalance impairments, which has already been deduced in (5.8), and is rewritten here:

$$x(n) = h_d(n) \otimes s(n) + h_m(n) \otimes s^*(n) \quad (6.9)$$

where $h_d(n)$ and $h_m(n)$ are the complex-valued direct filter and mirror filter, respectively. And hence $x(n)$ is a function of both $s(n)$ and $s^*(n)$, and can be written as:

$$\begin{aligned}
x(n) &= f_{IQ}(\mathbf{s}(n), \mathbf{s}^*(n)) \\
&= f_{IQ}(s(n), s(n-1) \cdots s(n-M_d+1), s^*(n), s^*(n-1) \cdots s^*(n-M_m+1))
\end{aligned} \tag{6.10}$$

where $f_{IQ}(\cdot)$ is a linear function, and M_d and M_m are the memory depths of the direct filter and mirror filter respectively. Note that M_d and M_m are equal to the length of the filters respectively.

Fig. 6-1 IQ imbalance and PA distortions at the wireless transmitter. The index n means the digital signal, while index t means the continuous analog signal after DAC.

Fig. 6-2 Diagram of TX IQ imbalance and PA impairments

The relationship between $x(n)$ and $y(n)$ is the PA nonlinear memory impairments, and can be written in the following form:

$$\begin{aligned} y(n) &= f_{PA}(\mathbf{x}(n)) \\ &= f_{PA}(x(n), x(n-1) \cdots x(n-M_{PA}+1)) \end{aligned} \quad (6.11)$$

where $f_{PA}(\cdot)$ is a nonlinear memory function, and M_{PA} is the memory depths of the power amplifier.

Hence the relationship between $s(n)$ and $y(n)$ has the following form:

$$\begin{aligned} y(n) &= f_{PA}(\mathbf{x}(n)) \\ &= f_{PA} \circ f_{IQ}(\mathbf{s}(n), \mathbf{s}^*(n)) \end{aligned} \quad (6.12)$$

where $f_{PA} \circ f_{IQ}$ denotes the cascaded relation of f_{IQ} and f_{PA} . If let $f = f_{PA} \circ f_{IQ}$, then (6.12) becomes

$$\begin{aligned} y(n) &= f(\mathbf{s}(n), \mathbf{s}^*(n)) \\ &= f(s(n), s(n-1) \cdots s(n-M+1), s^*(n), s^*(n-1) \cdots s^*(n-M+1)) \end{aligned} \quad (6.13)$$

where $f = f_{PA} \circ f_{IQ}$ is the unified function describing the cascaded TX IQ imbalance and PA impairments. The total memory depth is $M = M_{PA} + \max(M_d, M_m)$. Here $f(\cdot)$ is the prototype of the unified model.

From (6.13), the unified model is a function of two time-delayed inputs: $\mathbf{s}(n)$ and $\mathbf{s}^*(n)$, and is called *double-input* model. The structure is shown in Fig. 6-3. Since it is a single input model, the traditional MPM model [34] cannot model the joint PA and IQ imbalance.

It is important to note, however, that the models for power amplifiers are not double-input, because they only contain a single time-delayed input $\mathbf{s}(n)$. Hence the PA behavioral models, e.g., Volterra model and MPM model, cannot characterize the unified behavior of the TX IQ imbalance and PA impairments.

Fig. 6-3 First form of unified model for joint TX IQ imbalance and PA impairments

Equation (6.13) is the first form of the unified model, where the inputs are complex-valued signals $\mathbf{s}(n)$ and $\mathbf{s}^*(n)$. By simple transformation, the double-input unified model has another form. Let $s(n) = s_I(n) + js_Q(n)$ and $s^*(n) = s_I(n) - js_Q(n)$, (6.13) can also be written as:

$$\begin{aligned}
 y(n) &= f(\mathbf{s}_I(n) + j\mathbf{s}_Q(n), \mathbf{s}_I(n) - j\mathbf{s}_Q(n)) \\
 &= f'(\mathbf{s}_I(n), \mathbf{s}_Q(n)) \\
 &= f'(s_I(n), s_I(n-1) \cdots s_I(n-M), s_Q(n), s_Q(n-1) \cdots s_Q(n-M))
 \end{aligned} \tag{6.14}$$

where $s_I(n) = \Re\{s(n)\}$ and $s_Q(n) = \Im\{s(n)\}$ are the I-path component and Q-path component respectively. And (6.14) is the second form of double-input unified model, where the inputs are real-valued. The structure is shown in Fig. 6-4.

Fig. 6-4 Second form of unified model for joint TX IQ imbalance and PA impairments

By separating the I-path and Q-path components of the input signal, and treating them as two different input signals, the *degrees of freedom* is increased, given the same type of model. A simple example of a linear gain system in Fig. 6-5 is used to illustrate this. Denote the input signal $x(n) = a(n) + j \cdot b(n)$, and the complex gain $c = \alpha + j\beta$, then the output is

$$\begin{aligned}
 y(n) &= x(n) \cdot c \\
 &= (a(n) + jb(n)) \cdot (\alpha + j\beta) \\
 &= (\alpha a(n) - \beta b(n)) + j(\beta a(n) + \alpha b(n))
 \end{aligned} \tag{6.15}$$

Fig. 6-5 Linear Gain System

On the other hand, for the rectangular structure where the input signal is decouple into real and image parts $\hat{\mathbf{x}}(n) = [a(n) \ b(n)]^T$, which are treated as two different input signals, the linear gain is a two-scalar vector $\mathbf{c} = [c_1 \ c_2]^T$, and the output is:

$$\begin{aligned}
y(n) &= \begin{bmatrix} c_1 & c_2 \end{bmatrix} \cdot \begin{bmatrix} a(n) \\ b(n) \end{bmatrix} \\
&= \begin{bmatrix} \alpha_1 + j\beta_1 & \alpha_2 + j\beta_2 \end{bmatrix} \cdot \begin{bmatrix} a(n) \\ b(n) \end{bmatrix} \\
&= (\alpha_1 a(n) + \alpha_2 b(n)) + j(\beta_1 a(n) + \beta_2 b(n))
\end{aligned} \tag{6.16}$$

Comparing (6.15) and (6.16), it is obvious that in traditional single complex-input single complex-output linear gain system, it is implied that:

$$\begin{cases} \alpha_1 = \beta_2 \\ -\alpha_2 = \beta_1 \end{cases} \tag{6.17}$$

Hence by applying rectangular structure, the degrees of freedom is increased, i.e., more coefficients are available to model a real system with more accuracy and flexibility.

6.4.2 Rectangular Structured Focused Time-Delay Neural Network

The second form of unified model defined in (6.14) can be further decomposed into I component and Q component functions, as in the following form:

$$\begin{aligned}
y(n) &= f'(\mathbf{s}_I(n), \mathbf{s}_Q(n)) \\
&= f'_I(\mathbf{s}_I(n), \mathbf{s}_Q(n)) + j \cdot f'_Q(\mathbf{s}_I(n), \mathbf{s}_Q(n))
\end{aligned} \tag{6.18}$$

where it is obvious that $f'_I(\cdot)$ and $f'_Q(\cdot)$ are both real-valued I component and Q component functions. (6.18) is called *rectangular structure of the second form of unified model*. By transforming the complex-valued model in (6.14) into real-valued model, rectangular structure of the second form of unified model simplifies the implementation and identification of the model. The structure is shown in Fig. 6-6.

Fig. 6-6 Rectangular structure of the second form of unified model

Based on the above rectangular structure, a new unified model is proposed. The I component and Q component functions $f'_I(\cdot)$ and $f'_Q(\cdot)$ are implemented by Focused Time-Delay Neural Network (FTDNN) [178]. Focused Time-Delay Neural Network is one of the simplest neural network that incorporates memory effects. Since both $f'_I(\cdot)$ and $f'_Q(\cdot)$ are real-valued functions, the real-valued version of FTDNN is employed, i.e., the weights, biases, activation functions, and outputs are all real-valued. And from its rectangular structure, the proposed neural network is named *rectangular structured Focused Time-Delay Neural Network* (RSFTDNN) model.

The design and layout of RSFTDNN is shown in Fig. 6-7. The memory effects are modeled by using delayed inputs at the input layer, where z^{-1} means one sample delay. Hence the input layer is simply two tapped delay lines storing time-delayed inputs $[s_I(n), s_I(n-1) \cdots s_I(n-M)]$ and $[s_Q(n), s_Q(n-1) \cdots s_Q(n-M)]$, where M is the memory depth of the unified model.

The total number of layers is $L+1$; specifically, Layer 0 is the input layer, Layer 1 ~ Layer $L-1$ are the inner layers, and Layer L is the output layer. Note that the value of L can be varying according to different devices and requirements.

Moreover, let $\{N_i|0 \leq i \leq L\}$ denotes the number of neurons in each layer. The number of neurons in the input layer equals 2 times of the memory depth of the unified model, i.e., $N_0 = 2M$. The number of neurons in the output layer $N_L = 2$, which are the I and Q component outputs $y_I(n)$ and $y_Q(n)$. Meanwhile, different inner layers can have different number of neurons; and the value of N_i can be varying according to different devices and requirements.

The forward computation and the design of the neuron are discussed in the next section.

Fig. 6-7 RSFTDNN model, where z^{-1} means one sample delay. The number of inner layers, and the number of neurons of each inner layer can be varying according to different devices and requirements.

6.4.3 Neuron

The design of the neuron is shown in Fig. 6-8. Let $x_{i,(l+1)}(n)$ denote the output of the i^{th} neuron in the $l+1^{th}$ layer. To simplify the representation, the time sample index n is ignored. The number of neurons in the l^{th} layer is N_l . Then the outputs of $l+1^{th}$

layer are generated from the outputs of the l^{th} layer by the following mathematical relation:

$$\begin{cases} net_{i,(l+1)} = \sum_{j=1}^{N_l} \alpha_{ij,(l+1)} x_{j,(l)}(n) + \beta_{i,(l+1)} \\ x_{i,(l+1)}(n) = \Phi_{i,(l+1)}(net_{i,(l+1)}) \end{cases} \quad (6.19)$$

where $net_{i,(l+1)}$ is the weighted sum of the inputs, which are the outputs of the l^{th} layer, $\alpha_{ij,(l+1)}$ is the weights from the j^{th} neuron in the l^{th} layer to the i^{th} neuron in the $l+1^{th}$ layer, and $\beta_{i,(l+1)}$ is the bias for the i^{th} neuron in the $l+1^{th}$ layer. And $\Phi_{i,(l+1)}(\cdot)$ is called the *activation function* for the i^{th} neuron in the $l+1^{th}$ layer.

Fig. 6-8 Design of neuron: the i^{th} neuron in the $l+1^{th}$ layer

To further simplify the network, all the inner activation functions and the output layer activation function are the same, denoted as $\Phi(\cdot)$. Note that the AM/AM relationship is a ‘bend-down’ curve, i.e., the slope of the curve is decreasing with increasing input amplitude. To reflect this inherent ‘bend-down’ nature of power amplifier distortion, the activation function is chosen to be a *sigmoid* function, which also bends down for large input. The *Sigmoid* function is given by:

$$\Phi_{sgm}(x) = \frac{2c_1}{1 + e^{-c_2 \cdot x}} - c_1 \quad (6.20)$$

And its 1st-order derivative function is

$$\Phi'_{sgm}(x) = \frac{c_2}{2c_1} (c_1^2 - f_{sgm}^2(x)) \quad (6.21)$$

where parameter c_2 is the slope parameter, and c_1 defines the range of sigmoid function as $(0 \ c_1]$.

6.4.4 Identification

The identification method used is the back-propagation (BP) method [70, 179-183]. The basic idea is to minimize the error between the actual output and the RSFTDNN model output. The output error is denoted by $e(n) = y^0(n) - y(n)$, where $y^0(n)$ is the desired output, and $y(n)$ is the output of RSFTDNN model. And the complex-valued $e(n)$ has the form:

$$\begin{aligned} e(n) &= e_I(n) + j \cdot e_Q(n) \\ &= [y_I^0(n) - y_I(n)] + j \cdot [y_Q^0(n) - y_Q(n)] \end{aligned} \quad (6.22)$$

The cost function is in the form

$$\begin{aligned}
\varepsilon(n) &= |e(n)|^2 \\
&= |y^0(n) - y(n)|^2 \\
&= [y_I^0(n) - y_I(n)]^2 + [y_Q^0(n) - y_Q(n)]^2
\end{aligned} \tag{6.23}$$

The coefficient is updated by:

$$\begin{aligned}
\alpha_{ij,(l+1)}(n+1) &= \alpha_{ij,(l+1)}(n) + \frac{1}{2} \mu (-\nabla \varepsilon(n)) \\
&= \alpha_{ij,(l+1)}(n) + \mu \cdot \left[e_I(n) \cdot \frac{\partial y_I(n)}{\partial \alpha_{ij,(l+1)}(n)} + e_Q(n) \cdot \frac{\partial y_Q(n)}{\partial \alpha_{ij,(l+1)}(n)} \right]
\end{aligned} \tag{6.24}$$

and

$$\begin{aligned}
\beta_{i,(l+1)}(n+1) &= \beta_{i,(l+1)}(n) + \frac{1}{2} \mu (-\nabla \varepsilon(n)) \\
&= \beta_{i,(l+1)}(n) + \mu \cdot \left[e_I(n) \cdot \frac{\partial y_I(n)}{\partial \beta_{i,(l+1)}(n)} + e_Q(n) \cdot \frac{\partial y_Q(n)}{\partial \beta_{i,(l+1)}(n)} \right]
\end{aligned} \tag{6.25}$$

where μ is the step size to adjust the speed of the convergence. Use the chain rule of derivatives,

$$\begin{aligned}
\frac{\partial y_I(n)}{\partial \alpha_{ij,(l+1)}(n)} &= \frac{\partial y_I(n)}{\partial x_{i,(l+1)}(n)} \cdot \frac{\partial \Phi(\text{net}_{i,(l+1)}(n))}{\partial \text{net}_{i,(l+1)}(n)} \cdot \frac{\partial \text{net}_{i,(l+1)}(n)}{\partial \alpha_{ij,(l+1)}(n)} \\
&= \frac{\partial y_I(n)}{\partial x_{i,(l+1)}(n)} \cdot \Phi'(\text{net}_{i,(l+1)}(n)) \cdot x_{j,(l)}(n)
\end{aligned} \tag{6.26}$$

and

$$\begin{aligned}
\frac{\partial y_I(n)}{\partial \beta_{i,(l+1)}(n)} &= \frac{\partial y_I(n)}{\partial x_{j,(l+1)}(n)} \cdot \frac{\partial \Phi(\text{net}_{i,(l+1)}(n))}{\partial \text{net}_{i,(l+1)}(n)} \cdot \frac{\partial \text{net}_{i,(l+1)}(n)}{\partial \beta_{i,(l+1)}(n)} \\
&= \frac{\partial y_I(n)}{\partial x_{i,(l+1)}(n)} \cdot \Phi'(\text{net}_{i,(l+1)}(n))
\end{aligned} \tag{6.27}$$

The derivatives of the output with respect to any inner neuron can be computed in an iteratively back propagated way:

$$\begin{aligned}\frac{\partial^T y_I(n)}{\partial x_{j,(l)}(n)} &= \sum_{i=1}^{N_{(l+1)}} \frac{\partial y_I(n)}{\partial x_{ij,(l+1)}(n)} \cdot \frac{\partial^T x_{ij,(l+1)}(n)}{\partial x_{j,(l)}(n)} \\ &= \sum_{i=1}^{N_{(l+1)}} \frac{\partial y_I(n)}{\partial x_{ij,(l+1)}(n)} \cdot \Phi' \left(net_{ij,(l+1)}(n) \right) \cdot \alpha_{ij,(l+1)}(n)\end{aligned}\quad (6.28)$$

And

$$\begin{aligned}\frac{\partial^T y_Q(n)}{\partial x_{j,(l)}(n)} &= \sum_{i=1}^{N_{(l+1)}} \frac{\partial y_Q(n)}{\partial x_{ij,(l+1)}(n)} \cdot \frac{\partial^T x_{ij,(l+1)}(n)}{\partial x_{j,(l)}(n)} \\ &= \sum_{i=1}^{N_{(l+1)}} \frac{\partial y_Q(n)}{\partial x_{ij,(l+1)}(n)} \cdot \Phi' \left(net_{ij,(l+1)}(n) \right) \cdot \alpha_{ij,(l+1)}(n)\end{aligned}\quad (6.29)$$

(6.28) and (6.29) are the iterative update equation of the back propagated derivatives from the output of the network back to the inner neurons of the network. The complete description of back propagation method can be found in [70].

6.5 Experiment Results

6.5.1 Experiment setup

Fig. 6-9 shows the overall experimental setup. The DUT consists of the vector signal generator and the power amplifier (PA). Specifically, the signal generator contains TX IQ imbalance distortion, while power amplifier contains the nonlinear and memory PA distortion. Here the signal generator is Agilent ESG E4438C vector signal generator, which consists of the I-Path/Q-Path DACs, the reconstruction filters and the up-converters, as well as other analog devices. And the power amplifier is Mini-Circuits ZVE-8G Amplifier.

Fig. 6-9 Experiment devices setup diagram

The training signal is complex white uniformly-distributed random signal. It is to uniformly sample and explore the meaningful range of the mathematical model of the DUT, and give equal weight to different amplitude. The digital baseband training signal $s(n)$ is passed to the signal generator. Then the signal generator converts the digital baseband signal $s(n)$ to its corresponding RF signal by passing it through the I-Path/Q-Path DACs, the reconstruction filters and the up-converters, where the generated RF signal has carrier frequency of 2.4GHz and bandwidth of 16.6MHz. After that, the signal passes through the power amplifier, with the average output power of around 31 dBm that is also the 1dB gain compression point. As the Peak-to-Average Power Ratio (PAPR) of the input OFDM signal is above 10 dB, a significant part of the signal is working deeply in the nonlinear zone, and the peak amplitude reaches as deep as 10 dB higher than the 1-dB compression point. It is commonly considered as deep nonlinearity [25, 28, 184, 185].

In the receiver, the RF signal is first down-converted to an intermediate frequency signal of 100MHz by the down converter, and then received and digitized by the oscilloscope (Infiniium 54832D DSO) with sampling rate of 2Gsps. The digital signal is then passed to the personal computer. The digital IF signal is further down-converted to baseband by the digital demodulator, which prevents the signal from additional RX IQ imbalance distortion. After time alignment, the received baseband signal $z(n)$ and the transmitted signal $s(n)$ are used to identify the RSFTDNN model.

Totally 42,818 pairs of input-output data samples are used. Following the work of Doyle et al. [156] of using 50% of samples for model identification and 50% for model verification, the first 50% of the dataset (21,409 pairs) is used to train the RSFTDNN model, while the remaining 50% of the dataset (21,409 pairs) is used for testing. As a criterion, the normalized root mean square error (NRMSE) is used to evaluate the performance and accuracy of the model:

$$NRMSE = \sqrt{\frac{\sum_{n=1}^N |y(n) - z(n)|^2}{\sum_{n=1}^N |y(n)|^2}} \quad (6.30)$$

where N is the number of testing samples, $y(n)$ is the model output, and $z(n)$ is received baseband signal.

The number of inner layers of the proposed RSFTDNN model in the experiment is 2. For comparison, the performance of the memory polynomial model based dual input model [121, 122] is also evaluated and compared with the proposed model. Note that memory polynomial model [41] is a typical polynomial based model and one of the simplest version of the Volterra based model. Hence it is supposed to have the least

condition number problem and consume the least computation resources among polynomial based models. Its nonlinear order is set to be 5, the same as in [122].

6.5.2 RSFTDNN model of TX IQ Imbalance and PA Distortions

The joint behavior of the linear TX IQ imbalance distortion caused by the circuits inside the signal generator and the nonlinear memory distortions caused by power amplifier is modeled by the RSFTDNN model. The memory depth is 20, i.e., there are 20 delay taps at the input. And there are 2 inner layers, and 1 output layer. Each inner layer has 10 neurons, with *Sigmoid* activation function. There are 2 neurons in the output layer, and hence two outputs of the RSFTDNN model – one for I component, the other for Q component. After complex-addition as shown in Fig. 6-7, the final output is obtained.

The learning curve of the RSFTDNN model is shown in Fig. 6-10. The step size μ for update is 0.01. The learning curve converges after around 5,000 iterations. The fluctuation in the stable zone is due to the noise influence. Important physical parameters of the RF system are given in Table 6-1. The time-domain waveform comparison is demonstrated in Fig. 6-11. Here the RSFTDNN model output waveform is compared with the measured waveform in the real experiment. I component waveform and Q component waveform are compared separately. In Fig. 6-12, the AM/AM and AM/PM relationships obtained from the RSFTDNN model is compared with those obtained from measurement. The spectrum comparison is also shown in Fig. 6-13, where the spectrum of the RSFTDNN model output is compared with the real signal in the measurement. All figures demonstrate a good agreement between measured signal and the simulated signal by the RSFTDNN model, and justify the high accuracy of RSFTDNN model.

Fig. 6-10 Learning curve of RSFTDNN model

Parameters	Value
Real device output power level	31.37 dBm
RSFTDNN model output power level	31.42 dBm
Real device PAE	12.7%
RSFTDNN model PAE	12.8%
Real device gain	31.36 dB
RSFTDNN model gain	31.41 dB

Table 6-1 Physical parameters comparison between real device and RSFTDNN model

Fig. 6-11 Time domain waveform comparison

Fig. 6-12 AM/AM and AM/PM relations comparison

Fig. 6-13 Spectrum comparison, including the spectrum of the error between the measured and modeled results

The accuracy of the RSFTDNN model is further compared with that of the MPM-Dual-Input model. Here the memory depths of both models are swept from 1 to 20 taps, and the accuracies of both models are compared using the NRMSE defined in (6.30). As shown in Fig. 6-14, the accuracy of the RSFTDNN model does not change much when the memory depth is changed, but the accuracy of MPM-Dual-Input model drops dramatically with increased memory depth. The reason for the accuracy degradation of MPM-Dual-Input model is due to the *condition number* problem – the condition number is increasing very fast with memory depth, which is also illustrated in Fig. 6-14. Note that the maximum number of parameters of MPM-Dual-Input model is 252 for the case of 20 memory-tap depth. Comparing with the number of training samples (21,409 samples), the possibility of severe over-fitting problem [186] can be dismissed.

The resources required for the simulator between RSFTDNN model and MPM-Dual-Input model are compared in Fig. 6-15. 21,409 input samples are injected into the model, and the output is calculated and generated accordingly. The results show that RSFTDNN consume less memory space than MPM-Dual-Input and the gap increases with the model memory depth. For the case of 20 taps, RSFTDNN consumes less than 1/5 of the memory space that MPM-Dual-Input consumes. The simulation time for RSFTDNN remains relatively constant with increase in memory depth, whereas it increases 650% for the MPM-Dual-input when the memory depth increases from 1-tap to 20-taps. In absolute simulation time, the RSFTDNN has an advantage over MPM-Dual-Input for memory depth beyond 11 taps.

Fig. 6-14 Accuracy comparison between RSFTDNN model and MPM-Dual-Input model, and the condition number of MPM-Dual-Input model.

Fig. 6-15 Resources consumption comparison between RSFTDNN model and MPM-Dual-Input model: memory space consumption and simulation time consumption.

The comparison is based on a Hewlett-Packard Z210 workstation with Intel Xeon CPU

E31225 @3.10GHz and 4.00 GB RAM memory.

6.5.3 Effect of Increased IQ Imbalance

To simulate a low cost vector modulator that has poor IQ imbalance, additional IQ imbalance is added onto the inherent IQ imbalance in the signal generator. And in order to enhance the memory effects of the distortion, the additional IQ imbalance is implemented by four linear filters. The model is shown in Fig. 6-16, where the four low pass filters $h_{II}(n)$, $h_{IQ}(n)$, $h_{QI}(n)$, $h_{QQ}(n)$ are used to model the TX IQ imbalance distortion, with memory effects and cross-talking between I and Q-paths. Their coefficients are shown in Table 6-2. The overall experiment setup is shown in Fig. 6-16 and Fig. 6-17. The original transmitted signal $s(n)$ is firstly distorted by the IQ imbalance model. After that, the newly distorted signal $s'(n)$ is passed to the signal generator and the whole transmission

channel, where it further suffers from TX IQ imbalance and PA distortions in the physical devices.

Fig. 6-16 Additional IQ imbalance added in baseband

Fig. 6-17 Experiment devices setup diagram with additional TX IQ imbalance

$h_{II}(n)$	0.0021, 0.0734, 0.1546, 0.219, 0.2435, 0.219, 0.1546, 0.0734, 0.0021
$h_{IQ}(n)$	0.0006, 0.0026, 0.005, 0.0072, 0.0085, 0.0085, 0.0072, 0.005, 0.0026
$h_{QI}(n)$	0.0008, 0.0028, 0.0051, 0.0068, 0.0074, 0.0068, 0.0051, 0.0028, 0.0008
$h_{QQ}(n)$	0.0024, 0.0737, 0.155, 0.2195, 0.2441, 0.2195, 0.155, 0.0737, 0.0024

Table 6-2 Coefficients of the IQ imbalance model

The RSFTDNN model has the same structure as in the previous case. Fig. 6-18 shows the learning curve, which converges after 5,000 iterations. Again, the fluctuation in the stable zone is due to the noise influence. Important physical parameters of the RF system are given in Table 6-3. Time domain waveform of the RSFTDNN model is compared with the measured signals in Fig. 6-19, where each I and Q components are compared separately. The AM/AM and AM/PM relationships of the RSFTDNN model is compared with those obtained from the measured signals in Fig. 6-20. The spectrum comparison is shown in Fig. 6-21. All these comparisons show that the RSFTDNN model is very accurate, and the simulated output of the model is in good agreement with the measured signal.

Fig. 6-18 Learning curve of RSFTDNN model

The accuracy of the RSFTDNN model is also compared with that of the MPM-Dual-Input model, with the memory depths of both models being swept from 1 to 20 taps, as illustrated in Fig. 6-22. As in the previous case, the accuracy of RSFTDNN model does not change much when the memory depth is changed. On the contrary, owing to the fast-growing condition number, also illustrated in Fig. 6-22, the accuracy of the MPM-Dual-Input model is deteriorating very fast with increased memory depth. The resources required for the simulator of 21,409 input samples between RSFTDNN model and MPM-Dual-Input model are compared in Fig. 6-23.

Parameters	Value
Real device output power level	31.36 dBm
RSFTDNN model output power level	31.50 dBm
Real device PAE	12.7%
RSFTDNN model PAE	13.1%
Real device gain	31.35 dB
RSFTDNN model gain	31.49 dB

Table 6-3 Physical parameters comparison between real device and RSFTDNN model

Fig. 6-19 Time domain waveform comparison

Fig. 6-20 AM/AM and AM/PM relations comparison

Fig. 6-21 Spectrum comparison, including the spectrum of the error between the measured and modeled results

Fig. 6-22 Accuracy comparison between RSFTDNN model and MPM-Dual-Input model,
and the condition number of MPM-Dual-Input model

Fig. 6-23 Resources consumption comparison between RSFTDNN model and
MPM-Dual-Input model: memory space consumption and simulation time consumption.

The comparison is done based on the same computer as the previous experiment.

6.5.4 Test with OFDM Signals

The RSFTDNN models of Section 6.5.2 and Section 6.5.3 are further tested with OFDM signal. In both cases, the training signals are still the same complex white uniform distributed random signal, while the testing signals are 21,409 pairs of input-output 16-QAM OFDM data samples. The memory depth of RSFTDNN model is 20-tap.

The results using the OFDM testing signal are shown as: Fig. 6-24 and Fig. 6-25 show the time domain waveform comparison between the measured OFDM signal from the real experiment and the output of the RSFTDNN model. Fig. 6-26 and Fig. 6-27 compare the AM/AM and AM/PM relationships of the RSFTDNN model with that measured from the real experiment tested with the OFDM signal. The spectrums of the OFDM testing signal obtained from the real experiment are also compared with that obtained from RSFTDNN model in Fig. 6-28 and Fig. 6-29.

Fig. 6-24 Time domain waveform comparison for the case of Section 6.5.2, using OFDM testing signal

Fig. 6-25 Time domain waveform comparison for the case of Section 6.5.3, using OFDM testing signal

Fig. 6-26 AM/AM and AM/PM relation comparison for the case of Section 6.5.2, using OFDM testing signal

Fig. 6-27 AM/AM and AM/PM relation comparison for the case of Section 6.5.3, using
OFDM testing signal

Fig. 6-28 Spectrum comparison for the case of Section 6.5.2, using OFDM testing signal,
including the spectrum of the error between the measured and modeled results

Fig. 6-29 Spectrum comparison for the case of Section 6.5.3, using OFDM testing signal, including the spectrum of the error between the measured and modeled results

6.5.5 Structure Analysis and Comparison

The structure of RSFTDNN is shown in Fig. 6-7. It has two freedoms: (1) the number of inner layers $L-1$ (L equals inner layer number plus one output layer), (2) and the number of neurons of each inner layer $\{N_i | 1 \leq i \leq L-1\}$. In this section, these two freedoms are studied, and the performance of their different values is compared. The goal is to find the optimal layout of RSFTDNN structure.

Meanwhile, a TDNN model for joint power amplifier and IQ imbalance is proposed in [124]. It is real-valued TDNN model, with a structure of 2 inner layers, 1 input layer, and 1 output layer. The activation function is hyperbolic tangent. Its performance is also compared with the RSFTDNN of this thesis.

To facilitate comparison, the number of neurons is set equal for all inner layers, i.e., $\{N_1 = N_2 = \dots = N_{L-1} = N\}$. The memory depth at input layer is set to 20-taps. Then the value of N is swept for different layer number L , and the performance is compared.

Fig. 6-30 and Fig. 6-31 show the comparison for the case of Section 6.5.3, using Gaussian and OFDM testing signals, respectively. Label “TDNN2012mutual” refer to the model in [124]. Firstly, less number of inner layers achieves better accuracy – “L=2” achieves the best accuracy, with 5 ~ 6 dB improvement compared with “L=4”. Secondly, less number of inner layers has faster convergence rate – “L=2” converges at $N = 6$, while “L=4” converges at $N = 18$. As a result, “L=2” only needs 6 neurons at the inner layer to achieve the optimal performance. In the meantime, “TDNN2012mutual” has almost the same performance as “L=3”, because they both have the same layout of 2 inner layers.

From the comparison, it is clear that single-inner-layer is the optimal structure, as it achieves the best accuracy and least number of neurons. It is because that single-inner-layer structure is simpler, and has less number of coefficients to identify. And according to the *universal approximation theory* [70, 187, 188], single-inner-layer neural network with *sigmoid* activation function can approximate any continuous function with arbitrary accuracy.

Fig. 6-30 Comparison of different layouts: number of inner layers, and number of neurons of each inner layer. It is for the case of Section 6.5.3, using Gaussian testing signal.

Fig. 6-31 Comparison of different layouts: number of inner layers, and number of neurons of each inner layer. It is for the case of Section 6.5.3, using OFDM testing signal.

6.6 Conclusion

RSFTDNN model is proposed to unify and characterize the behavior of the TX IQ imbalance and power amplifier. It saves computation resources in system-level simulation, provides unified view of the transmitter system, and can be conveniently implemented in experiment. The RSFTDNN model can accurately characterize both mild and deep nonlinear system. Its structure is simple to implement, and is stable because there is no feedback loop. The model is tested on a transmitter with IQ imbalance and the power amplifier operated into gain compression mode.

Results show very good agreement between measured and modeled in time-domain waveform, the AM/AM and AM/PM relationships, and the transmitted spectrum. Compared with the dual input memory polynomial model, the RSFTDNN model uses fewer resources and is more accurate. The RSFTDNN model also shows good performance for OFDM signals.

Chapter 7 Conclusion and Future Works

7.1 Conclusion

This thesis is mainly focused on the behavioral models of the most important analog impairments in the modern wireless communication systems – power amplifier nonlinear memory distortion and transmitter IQ imbalance distortion. The research is concentrated in several key areas: design an accurate, sufficient, and flexible experiment test bed to facilitate the research on behavioral models, propose appropriate and accurate behavioral models to characterize the behaviors of power amplifier and IQ imbalance, apply accurate and robust numerical methods to extract the coefficients of the behavioral models under noise influence, and use the behavioral models to simulate or compensate the analog impairments in the wireless transmitter systems.

Chapter 2 introduces the background information on the distortions of TX IQ Imbalance and power amplifier. It discusses the subjects of system-level simulation and digital predistortion, and highlights the importance of the behavioral models in these applications. A comprehensive review on the behavioral models of TX IQ imbalance and power amplifier is conducted from which the drawbacks of current behavioral models are deduced. It highlights the necessities to find a better model that is accurate, simply to implement, easy to identify, and without generating extra impairments.

An adequate test bed is the fundamental and prerequisite condition for the research on behavioral models of the analog impairments in wireless communication systems. The key

issue here is that it should accurately recover the envelopes of the input and output signals of the device under test. In Chapter 3, a new accurate and flexible test bed is proposed. It is based on the common lab equipment, so that the cost of buying any specific equipment is saved. It implements competent signal processing methods, so that the various extra impairments through the test bed channel are well compensated. It is designed to achieve both high SNR and high linearity. It also provides the flexibility for any changes on the hardware and software. And it provides the facilities and metrics to evaluate the performance of the behavioral model, such as error vector magnitude and spectrum analysis.

To compensate the power amplifier nonlinear memory distortion, a new predistortion scheme is proposed in Chapter 4. The proposed model is based on the physical characteristics of the power amplifier – the AM/AM and AM/PM relationships. To be more specific, the predistorter is to separately characterize the inverse AM/AM and inverse AM/PM functions. And each function is a real-valued memory polynomial model. As a result, the proposed predistorter is single input double output structure with real coefficients. Unlike the classical memory polynomial model that only utilizes the *diagonal* elements of the Volterra model, the proposed model has both the *diagonal* elements and many *cross-term* elements. Since it is based on the PA physical characteristics and has much more elements, the proposed model has a much better performance than the classical memory polynomial model. At the same time, the proposed model is still very simple – it only contains two separate memory polynomial models. Hence it is easy to identify and implement.

To compensate TX IQ imbalance distortion of significant memory effects, a new transmitter predistortion scheme working in frequency domain is proposed in Chapter 5. It

is a frequency-domain inverse behavioral model that is used to compensate TX IQ imbalance. The basic idea is to transform the problem into a matrix problem, and then solve it by robust numerical methods. Comparing with the time-domain predistortion methods, the proposed frequency-domain predistorter can compensate deep memory effects, and it does not generate any extra inter-symbol interference in the signal. Since OFDM systems have inherent IFFT function at the transmitter, the proposed predistortion scheme does not need additional Fourier transform module. Hence it is simple, and can be easily implemented either in digital circuits or in microprocessors. The proposed predistortion scheme achieves good EVM performance in the experiment.

In Chapter 6, a new behavioral model is proposed to unify the characterization of the two analog impairments – TX IQ imbalance and power amplifier – and merge them into one single module. It is named rectangular structured Focused Time-Delay Neural Network (RSFTDNN) model. It can accurately model both mild and deep nonlinear devices. It provides the unified view of the analog impairments in the transmitter. It is designed to have double inputs, so as to characterize the IQ imbalance behavior. It is also designed to have the “bend-down” activation function, so as to characterize the PA nonlinear behavior. And with a non-feedback structure, it is simple and stable. The experiment results demonstrate that the proposed model is significantly more accurate than the classical dual-input memory polynomial model.

Meanwhile, information on the numerical issues discussed in the thesis is provided in the appendices, including singular value decomposition, traditional least squares problem, total least squares problem, and data least squares problem.

7.2 Future Works

The advantage of a polynomial based model is that it is in linear relationship with respect to its coefficients. Hence, various linear numerical methods can be used to identify the model. However, as discussed in Section 5.3 and Section 6.3, polynomial based models have serious problems in numerical issues – their accuracy is very sensitive to noise. The final error of a model is the sum of propagated data error and computational error (Please refer to Section 5.5.1). Although their computational error can be greatly reduced by some robust numerical methods, the propagated data error of polynomial based models cannot be alleviated by current methods. The PA predistortion scheme in Chapter 4 is based on memory polynomial model, which is one of the simplest forms of polynomial based models. Hence the model in Chapter 4 should have the least propagated-data-error, and successfully compensate the distortion of a power amplifier of mild nonlinearity and short memory effects. However, as the nonlinearity and memory effects increase with bandwidth and data rate of the signal, memory polynomial model will have very big condition number and propagated data error. Future work can investigate on the numerical issues of polynomial based models, and find a way to circumvent the condition number problem and reduce the propagated data error.

The frequency-domain predistortion scheme in Chapter 5 has the advantage that it does not utilize any additional memory compensator, and hence will not generate ISI impairment. Currently, our research work is only focused on using it to compensate TX IQ imbalance. At the same time, frequency-domain predistortion is equally promising for compensation of power amplifier distortion, or even to compensate the distortions of joint PA and TX IQ imbalance. Hence the future work may investigate on these subjects.

The RSFTDNN model in Chapter 6 does not have the same numerical problem as polynomial based models, and hence has the potential to model the devices of deep nonlinearity and long memory effects. It can be easily implemented in any computers or microprocessors to achieve most of the tasks. However, from cost-area-consideration, it is sometimes better to implement it in an integrated circuit (IC) or a chip. In that case, the activation functions of the neural network should be easy to be implemented by circuits. Hence the future work should find the activation functions that can be easily implemented in circuits, while without compromising the performance.

Finally, in this thesis, the research on behavioral models of power amplifier and IQ imbalance is based on modern wireless data communication systems for mobile phones and other data terminals. It is to cater for the communication systems of high-speed data in urban areas. At the same time, there are other kinds of communication systems, e.g., satellite navigation systems, fiber-optic communication systems, and power line communication systems, etc. These systems have different requirements on signal parameters such as bandwidth, output power level, and data rates, etc. As a result, the power amplifiers and modulators may show different performance and characteristics. Hence future work can investigate on how to apply and modify the current behavioral models to cater for the requirements of other systems.

Author's Publications

- [1] **K. Fu**, C. L. Law, C. Gu, T. T. Thein, and M. T. Kyaw, "A new power amplifier predistortion architecture based on memory polynomial model," in *7th International Conference on Information, Communications and Signal Processing*, pp. 1-5, 2009.
- [2] **K. Fu**, C. L. Law, and T. T. Thein, "Novel Neural Network Model of Power Amplifier Plus IQ Imbalances," *Progress In Electromagnetics Research B*, vol. 46, pp. 177-192, 2013.
- [3] **K. Fu**, C. L. Law, and T. T. Thein, "Test Bed for Power Amplifier Behavioral Characterization and Modeling," *Measurement*, vol. 46, pp. 2735--2745, 2013.
- [4] T. T. Thein, C. L. Law, and **K. Fu**, "Frequency Domain Dynamic Thermal Analysis in GaAs HBT for Power Amplifier Applications," *Progress In Electromagnetics Research*, vol. 118, pp. 71-87, 2011.
- [5] T. T. Thein, C. L. Law, and **K. Fu**, "250 MHz to 30 GHz, Unilateral Circuitmodel for InGap/GaAs HBT," *Progress In Electromagnetics Research C*, vol. 26, pp. 1-12, 2012.

Bibliography

- [1] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Deembedding static nonlinearities and accurately identifying and modeling memory effects in wide-band RF transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 3578-3587, 2005.
- [2] V. P. Papantoniou and T. D. Xenos, "Development and electromagnetic characterization of adaptable open-architecture WLAN systems," *Progress In Electromagnetics Research*, vol. 91, pp. 229-242, 2009.
- [3] I. W. Group, "IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems," *IEEE Standard*, pp. 802.16-2004, 2004.
- [4] S. Sesia, I. Toufik, and M. Baker, *LTE-the UMTS long term evolution: from theory to practice*: Wiley, 2009.
- [5] T. I. o. E. a. E. Engineers, "IEEE Standard for Telecommunications and Information Exchange Between Systems - LAN/MAN Specific Requirements - Part 11: Wireless Medium Access Control (MAC) and physical layer (PHY) specifications: High Speed Physical Layer in the 5 GHz band," ed: IEEE-SA Standards Board, 1999.
- [6] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*: Wiley, 2012.
- [7] T. C. W. Schenk, P. F. M. Smulders, and E. R. Fledderus, "Estimation and compensation of TX and RX IQ imbalance in OFDM-based MIMO systems," *IEEE Radio and Wireless Symposium*, pp. 215-218, 2006.
- [8] Y. Yoshida, K. Hayashi, H. Sakai, and W. Bocquet, "Analysis and compensation of transmitter IQ imbalances in OFDMA and SC-FDMA systems," *IEEE Transactions on Signal Processing*, vol. 57, pp. 3119-3129, 2009.
- [9] B. Narasimhan, D. Wang, S. Narayanan, H. Minn, and N. Al-Dhahir, "Digital compensation of frequency-dependent joint Tx/Rx I/Q imbalance in OFDM systems under high mobility," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, pp. 405-417, 2009.
- [10] D. White, "Johann Christian Doppler and his effect--a brief history," *Ultrasound in medicine & biology*, vol. 8, p. 583, 1982.
- [11] T. Jiang and Y. Wu, "An overview: Peak-to-average power ratio reduction techniques for OFDM signals," *IEEE Transactions on Broadcasting*, vol. 54, pp. 257-268, 2008.
- [12] F. H. Raab, P. Asbeck, S. Cripps, P. B. Kenington, Z. B. Popovic, N. Potheary, *et al.*, "RF and microwave power amplifier and transmitter technologies," *High Frequency Electronics*, vol. 2, pp. 22-36, 2003.
- [13] T. C. W. Schenk, P. F. M. Smulders, and E. R. Fledderus, "Impact of nonlinearities in multiple-antenna OFDM transceivers," *Symposium on Communications and Vehicular Technology*, pp. 53-56, 2006.
- [14] J. G. Proakis, *Digital communications*, 5th ed.: McGraw-Hill, 2007.
- [15] A. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Transactions on Communications*, vol. 29, pp. 1715-1720, 1981.
- [16] J. Fuenzalida, O. Shimbo, and W. Cook, "Time-domain analysis of intermodulation effects caused by nonlinear amplifiers," *COMSAT Technical Review*, vol. 3, pp. 89-143, 1973.
- [17] M. O'Droma, "Dynamic range and other fundamentals of the complex Bessel function series approximation model for memoryless nonlinear devices," *IEEE Transactions on Communications*, vol. 37, pp. 397-398, 1989.
- [18] P. Hetrakul and D. Taylor, "Nonlinear quadrature model for a travelling-wave-tube-type amplifier," *Electronics Letters*, vol. 11, p. 50, 1975.
- [19] A. Berman and C. Mahle, "Nonlinear phase shift in traveling-wave tubes as applied to multiple access communications satellites," *IEEE Transactions on Communication Technology*, vol. 18, pp. 37-48, 1970.
- [20] M. Schetzen, "Nonlinear system modeling based on the Wiener theory," *Proceedings of the IEEE*, vol. 69, pp. 1557-1573, 1981.

- [21] V. Volterra, "Theory of Functionals," ed: Blackie, 1930.
- [22] J. S. Bendat, *Nonlinear system techniques and applications*: John Wiley & Sons, Inc., 1998.
- [23] 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, pp. 348-359, 2006.
- [24] T. Ogunfunmi, *Adaptive nonlinear system identification: the Volterra and Wiener model approaches*: Springer, 2007.
- [25] W. J. Rugh, *Nonlinear system theory*: Johns Hopkins University Press, 1981.
- [26] V. J. Mathews and G. L. Sicuranza, *Polynomial signal processing*: Wiley New York, 2000.
- [27] 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, pp. 1150-1163, 2005.
- [28] D. Schreurs, M. O'Droma, A. A. Goacher, and M. Gadringer, *RF power amplifier behavioral modeling*: Cambridge University Press, 2008.
- [29] P. L. Gilabert, G. Montoro, and E. Bertran, "A methodology to model and predistort short-term memory nonlinearities in power amplifiers," *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*, pp. 142-145, 2006.
- [30] D. Silveira, M. Gadringer, H. Arthaber, M. Mayer, and G. Magerl, "Modeling, analysis and classification of a PA based on identified Volterra kernels," *European Gallium Arsenide and Other Semiconductor Application Symposium*, pp. 405-408, 2005.
- [31] A. Zhu and T. J. Brazil, "An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers," *IEEE Wireless and Microwave Technology Conference*, pp. 1-5, 2006.
- [32] R. Blum and M. Jeruchim, "Modeling nonlinear amplifiers for communication simulation," *IEEE International Conference on Communications 'World Prosperity Through Communications'*, pp. 1468-1472 vol. 3, 1989.
- [33] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of communication systems: modeling, methodology, and techniques*: Springer, 2000.
- [34] 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, pp. 3852-3860, 2006.
- [35] W. Bosch and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, pp. 1885-1890, 1989.
- [36] J. H. K. Vuolevi, T. Rahkonen, and J. P. A. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 1383-1389, 2001.
- [37] J. C. Pedro, N. B. Carvalho, and P. M. Lavrador, "Modeling nonlinear behavior of band-pass memoryless and dynamic systems," *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 2133-2136, 2003.
- [38] J. Vuolevi and T. Rahkonen, *Distortion in RF power amplifiers*: Artech house publishers, 2003.
- [39] P. Asbeck, H. Kobayashi, M. Iwamoto, G. Hanington, S. Nam, and L. Larson, "Augmented behavioral characterization for modeling the nonlinear response of power amplifiers," *IEEE MTT-S International Microwave Symposium Digest*, vol. 1, pp. 135-138, 2002.
- [40] P. Draxler, I. Langmore, T. Hung, and P. Asbeck, "Time domain characterization of power amplifiers with memory effects," *IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 803-806, 2003.
- [41] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, pp. 1417-1418, 2001.
- [42] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 4323-4332, 2006.
- [43] A. Zhu, J. Dooley, and T. J. Brazil, "Simplified Volterra series based behavioral modeling of RF power amplifiers using deviation-reduction," *IEEE MTT-S International Microwave Symposium Digest*, pp. 1113-1116, 2006.

- [44] 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," *IEEE/MTT-S International Microwave Symposium*, pp. 1441-1444, 2007.
- [45] N. Safari, T. Roste, P. Fedorenko, and J. S. Kenney, "An approximation of Volterra series using delay envelopes, applied to digital predistortion of RF power amplifiers with memory effects," *IEEE Microwave and Wireless Components Letters*, vol. 18, pp. 115-117, 2008.
- [46] M. Schetzen, *The Volterra and Wiener theories of nonlinear systems*: John Wiley & Sons, 1980.
- [47] R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 1468-1479, 2004.
- [48] D. Ronnow and M. Isaksson, "Digital predistortion of radio frequency power amplifiers using Kautz-Volterra model," *Electronics Letters*, vol. 42, pp. 780-782, 2006.
- [49] M. Isaksson and D. Ronnow, "A Kautz-Volterra behavioral model for RF power amplifiers," *IEEE MTT-S International Microwave Symposium Digest*, pp. 485-488, 2006.
- [50] A. Zhu and T. J. Brazil, "RF power amplifier behavioral modeling using Volterra expansion with Laguerre functions," *IEEE MTT-S International Microwave Symposium Digest*, p. 4, 2005.
- [51] Q. Zheng and E. Zafiriou, "Nonlinear system identification for control using Volterra-Laguerre expansion," *Proceedings of the American Control Conference*, vol. 3, pp. 2195-2199 1995.
- [52] C. Siviero, P. M. Lavrador, and J. C. Pedro, "A Frequency Domain Extraction Procedure of Low-Pass Equivalent Behavioural Models of Microwave PAs," *The 1st European Microwave Integrated Circuits Conference*, pp. 253-256, 2006.
- [53] M. C. Chiu, C. H. Zeng, and M. C. Liu, "Predistorter based on frequency domain estimation for compensation of nonlinear distortion in OFDM systems," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 882-892, 2008.
- [54] S. Kim and E. Powers, "Orthogonalised frequency domain Volterra model for non-Gaussian inputs," *IEE Proceedings of Radar and Signal Processing*, vol. 140, pp. 402-409, 1993.
- [55] G. Favier, A. Y. Kibangou, and A. Khouaja, "Nonlinear system modelling by means of Volterra models. approaches for the parametric complexity reduction," in *Invited paper in Symposium Techniques Avancées et Stratégies Innovantes en Modélisation et Commande Robuste des processus industriels*, 2004.
- [56] H. Ku and J. S. Kenney, "Behavioral modeling of nonlinear RF power amplifiers considering memory effects," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 2495-2504, 2003.
- [57] 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, pp. 563-565, 2004.
- [58] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, *et al.*, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Transactions on Communications*, vol. 52, pp. 159-165, 2004.
- [59] A. Ahmed, E. Srinidhi, and G. Kompa, "Neural network and memory polynomial methodologies for PA modeling," *7th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services*, vol. 2, pp. 393-396, 2005.
- [60] T. Liu, S. Boumaiza, and F. M. Ghannouchi, "Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, pp. 1025-1033, 2004.
- [61] Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial neural networks for RF and microwave design-from theory to practice," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 1339-1350, 2003.
- [62] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real-valued focused time-delay line neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 95-104, 2010.

- [63] Z. Huang, W. Chen, Z. Feng, and F. M. Ghannouchi, "Forward behavioral modeling of concurrent dual-band power amplifiers using extended real valued time delay neural networks," *International Conference on Microwave and Millimeter Wave Technology*, vol. 5, pp. 1-4, 2012.
- [64] J. Pochmara, "Improving compensation of nonlinear distortions in OFDM system using recurrent neural network with conjugate gradient algorithm," *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 180-185 Vol. 1, 2004.
- [65] D. Luongvinh and Y. Kwon, "A fully recurrent neural network-based model for predicting spectral regrowth of 3G handset power amplifiers with memory effects," *IEEE Microwave and Wireless Components Letters*, vol. 16, pp. 621-623, 2006.
- [66] Y. Cao, X. Chen, and G. Wang, "Dynamic behavioral modeling of nonlinear microwave devices using real-time recurrent neural network," *IEEE Transactions on Electron Devices*, vol. 56, pp. 1020-1026, 2009.
- [67] M. Isaksson, D. Wisell, and D. Ronnow, "Wide-band dynamic modeling of power amplifiers using radial-basis function neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 3422-3428, 2005.
- [68] M. Rawat and F. M. Ghannouchi, "Distributed spatiotemporal neural network for nonlinear dynamic transmitter modeling and adaptive digital predistortion," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, pp. 595-608, 2012.
- [69] A. Jarndal, S. Pillai, H. Abdulqader, and F. M. Ghannouchi, "A genetic neural network modeling of GaN HEMTs for RF power amplifiers design," *International Conference on Microelectronics*, pp. 1-6, 2011.
- [70] S. Haykin, *Adaptive Filter Theory*, 3rd ed.: Prentice Hall, 1996.
- [71] S. R. D. Paulo, *Adaptive filtering: algorithms and practical implementation*, 2nd ed.: Publishing House of Electronics Industry, 2002.
- [72] J. Bouras, S. Bouras, T. Georgantas, N. Haralabidis, G. Kamoulakos, C. Kapnistis, *et al.*, "A digitally calibrated 5.15-5.825 GHz transceiver for 802.11 a wireless LANs in 0.18 μ m CMOS," *IEEE International Solid-State Circuits Conference*, pp. 352-498, 2003.
- [73] J. P. F. Glas, "Digital I/Q imbalance compensation in a low-IF receiver," *IEEE Global Telecommunications Conference*, vol. 3, pp. 1461-1466, 1998.
- [74] R. A. Green, R. Anderson-Sprecher, and J. W. Pierre, "Quadrature receiver mismatch calibration," *IEEE Transactions on Signal Processing*, vol. 47, pp. 3130-3133, 1999.
- [75] H. G. Park, C. Park, H. Oh, and M. G. Kyeong, "RF gain/phase and I/Q imbalance error correction technique for multi-channel array antenna systems," *IEEE VTS 53rd Vehicular Technology Conference*, vol. 1, pp. 175-179, 2001.
- [76] X. Zhang, L. E. Larson, P. M. Asbeck, and P. Nanawa, "Gain/phase imbalance-minimization techniques for LINC transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 2507-2516, 2001.
- [77] C. Hsu and W. Sheen, "Joint Calibration of Transmitter and Receiver Impairments in Direct-Conversion Radio Architecture," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 832-841, 2012.
- [78] R. A. Green, "An optimized multi-tone calibration signal for quadrature receiver communication systems," *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing*, pp. 664-667, 2000.
- [79] A. Tarighat, R. Bagheri, and A. H. Sayed, "Compensation schemes and performance analysis of IQ imbalances in OFDM receivers," *IEEE Transactions on Signal Processing*, vol. 53, pp. 3257-3268, 2005.
- [80] S. Fouladifard and H. Shafiee, "Frequency offset estimation in OFDM systems in the presence of IQ imbalance," *IEEE International Conference on Communications*, vol. 3, pp. 2071-2075, 2003.
- [81] H. Lin, X. Zhu, and K. Yamashita, "Low-complexity pilot-aided compensation for carrier frequency offset and I/Q imbalance," *IEEE Transactions on Communications*, vol. 58, pp. 448-452, 2010.

- [82] Y. C. Pan and S. M. Phoong, "A Time-Domain Joint Estimation Algorithm for CFO and I/Q Imbalance in Wideband Direct-Conversion Receivers," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 2353-2361, 2012.
- [83] S. Fouladifard and H. Shafiee, "On adaptive cancellation of IQ mismatch in OFDM receivers," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. IV-564-7, 2003.
- [84] S. Fouladifard and H. Shafiee, "A new technique for estimation and compensation of IQ imbalance in OFDM receivers," *The 8th International Conference on Communication Systems*, vol. 1, pp. 224-228, 2002.
- [85] F. Horlin, A. Bourdoux, and L. Van der Perre, "Low-complexity EM-based joint acquisition of the carrier frequency offset and IQ imbalance," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 2212-2220, 2008.
- [86] F. Yan, W. P. Zhu, and M. Omair Ahmad, "Carrier frequency offset estimation for OFDM systems with I/Q imbalance," *47th IEEE Midwest Symposium on Circuits and Systems*, vol. 2, pp. II-633-II-636, 2004.
- [87] J. Tubbax, A. Fort, L. Van der Perre, S. Donnay, M. Engels, M. Moonen, *et al.*, "Joint compensation of IQ imbalance and frequency offset in OFDM systems," *IEEE Global Telecommunications Conference*, vol. 4, pp. 2365-2369, 2003.
- [88] J. Y. Yu, M. F. Sun, T. Y. Hsu, and C. Y. Lee, "A novel technique for I/Q imbalance and CFO compensation in OFDM systems," *IEEE International Symposium on Circuits and Systems*, pp. 6030-6033, 2005.
- [89] M. Valkama, M. Renfors, and V. Koivunen, "Compensation of frequency-selective I/Q imbalances in wideband receivers: models and algorithms," in *IEEE Third Workshop on Signal Processing Advances in Wireless Communications*, 2001.
- [90] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," *IEEE Transactions on Signal Processing*, vol. 49, pp. 2335-2344, 2001.
- [91] M. Valkama, K. Salminen, and M. Renfors, "Digital I/Q imbalance compensation in low-IF receivers: Principles and practice," *14th International Conference on Digital Signal Processing*, vol. 2, pp. 1179-1182, 2002.
- [92] S. C. Douglas and S. Haykin, "On the relationship between blind deconvolution and blind source separation," *31st Asilomar Conference on Signals, Systems & Computers*, vol. 2, pp. 1591-1595, 1997.
- [93] P. Rykaczewski, J. Brakensiek, and F. K. Jondral, "Decision directed methods of I/Q imbalance compensation in OFDM systems," *IEEE 60th Vehicular Technology Conference*, vol. 1, pp. 484-487, 2004.
- [94] A. Tarighat and A. H. Sayed, "OFDM systems with both transmitter and receiver IQ imbalances," *IEEE 6th Workshop on Signal Processing Advances in Wireless Communications*, pp. 735-739, 2005.
- [95] A. Tarighat and A. H. Sayed, "Joint compensation of transmitter and receiver impairments in OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 6, pp. 240-247, 2007.
- [96] E. Manasseh, S. Ohno, and M. Nakamoto, "Training Symbol Design for Channel Estimation and IQ Imbalance Compensation in OFDM Systems," *IEEE 75th Vehicular Technology Conference (VTC Spring)*, pp. 1-5, 2012.
- [97] L. Yan, S. Feng, S. Xiajie, F. Haiqiang, B. Stevan, and L. Jinhui, "Improved joint compensation scheme of transmitter and receiver IQ imbalances in OFDM systems," *IEEE International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1-4, 2011.
- [98] J. Feigin and D. Brady, "Joint transmitter/receiver I/Q imbalance compensation for direct conversion OFDM in packet-switched multipath environments," *IEEE Transactions on Signal Processing*, vol. 57, pp. 4588-4593, 2009.
- [99] R. M. Rao and B. Daneshrad, "I/Q mismatch cancellation for MIMO-OFDM systems," *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, pp. 2710-2714, 2004.
- [100] H. Kamata, K. Sakaguchi, and K. Araki, "An effective IQ imbalance compensation scheme for MIMO-OFDM communication system," *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 1, pp. 181-185, 2005.

- [101] D. Tandur and M. Moonen, "Decoupled compensation of IQ imbalance in MIMO OFDM systems," *Signal Processing*, vol. 91, pp. 1194-1209, 2011.
- [102] B. Narasimhan, S. Narayanan, H. Minn, and N. Al-Dhahir, "Reduced-complexity baseband compensation of joint Tx/Rx I/Q imbalance in mobile MIMO-OFDM," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 1720-1728, 2010.
- [103] H. Minn and D. Munoz, "Pilot designs for channel estimation of MIMO OFDM systems with frequency-dependent I/Q imbalances," *IEEE Transactions on Communications*, vol. 58, pp. 2252-2264, 2010.
- [104] J. Gao, X. Zhu, H. Lin, and A. K. Nandi, "Independent component analysis based semi-blind I/Q imbalance compensation for MIMO OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 9, pp. 914-920, 2010.
- [105] L. He, S. Ma, Y. C. Wu, and T. S. Ng, "IQ imbalance compensation: A semi-blind method for OFDM systems in fast fading channels," *IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*, pp. 362-365, 2010.
- [106] M. Beheshti, M. J. Omid, and A. M. Doost-Hoseini, "Joint Compensation of Transmitter and Receiver IQ Imbalance for MIMO-OFDM Over Doubly Selective Channels," *Wireless Personal Communications*, pp. 1-23, 2012.
- [107] I. Barhumi and M. Moonen, "IQ-imbalance compensation for OFDM in the presence of IBI and carrier-frequency offset," *IEEE Transactions on Signal Processing*, vol. 55, pp. 256-266, 2007.
- [108] F. Dwiya, R. Karlina, and T. Adiono, "Algorithm and hardware architecture of high accuracy IQ imbalance estimation and compensation for OFDM system," *International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, pp. 669-673, 2012.
- [109] H. Miyashita, M. Inamori, Y. Sanada, and T. Ide, "IQ Imbalance Estimation Scheme with Inter-carrier Interference Self-Cancellation Pilot Symbols in OFDM Direct Conversion Receivers," *IEEE 75th Vehicular Technology Conference (VTC Spring)*, pp. 1-5, 2012.
- [110] M. Windisch and G. Fettweis, "Adaptive I/Q imbalance compensation in low-IF transmitter architectures," *IEEE 60th Vehicular Technology Conference*, vol. 3, pp. 2096-2100, 2004.
- [111] D. Tandur and M. Moonen, "Joint adaptive compensation of transmitter and receiver IQ imbalance under carrier frequency offset in OFDM-based systems," *IEEE Transactions on Signal Processing*, vol. 55, pp. 5246-5252, 2007.
- [112] X. Huang and M. Caron, "Gain/phase imbalance and DC offset compensation in quadrature modulators," *IEEE International Symposium on Circuits and Systems*, vol. 4, pp. IV-811-IV-814 vol. 4, 2002.
- [113] C. C. Gaudes, M. Valkama, M. Renfors, and J. Ajanki, "Fast frequency synthesizer concept based on digital tuning and I/Q signal processing," *14th International Conference on Digital Signal Processing*, vol. 2, pp. 1317-1320, 2002.
- [114] M. Petermann, M. Niemiec, D. Wubben, K. D. Kammeyer, M. Stefer, and M. Schneider, "Iterative pre-compensation of I/Q imbalances in an adaptive 2.4 GHz MIMO-OFDM system," *International ITG Workshop on Smart Antennas (WSA)*, pp. 197-204, 2010.
- [115] A. G. K. C. Lim, V. Sreeram, and G. Q. Wang, "Digital compensation in IQ modulators using adaptive FIR filters," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 1809-1817, 2004.
- [116] J. Tuthill and A. Cantoni, "Optimum precompensation filters for IQ modulation systems," *IEEE Transactions on Communications*, vol. 47, pp. 1466-1468, 1999.
- [117] L. Ding, Z. Ma, D. R. Morgan, M. Zierdt, and T. Zhou, "Compensation of frequency-dependent gain/phase imbalance in predistortion linearization systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, pp. 390-397, 2008.
- [118] Y. Zou, M. Valkama, and M. Renfors, "Pilot-based compensation of frequency-selective I/Q imbalances in direct-conversion OFDM transmitters," *IEEE 68th Vehicular Technology Conference*, pp. 1--5, 2008.

- [119] R. K. McPherson, "Frequency-selective I/Q imbalance compensation for OFDM transmitters using online frequency-domain adaptive predistortion," *Military Communications Conference*, pp. 532--537, 2011.
- [120] L. Anttila, P. Handel, and M. Valkama, "Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 730-739, 2010.
- [121] H. Cao, S. Tehrani, C. Fager, T. Eriksson, and H. Zirath, "I/Q imbalance compensation using a nonlinear modeling approach," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, pp. 513-518, 2009.
- [122] H. Cao, A. S. Tehrani, C. Fager, T. Eriksson, and H. Zirath, "Dual-input nonlinear modeling for I/Q modulator distortion compensation," *IEEE Radio and Wireless Symposium*, pp. 39-42, 2009.
- [123] K. M. Gharaibeh, O. Al - Zoubi, and A. Alzayed, "Adaptive predistortion using threshold decomposition - based piecewise linear modeling," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 21, pp. 145-156, 2011.
- [124] M. Rawat and F. M. Ghannouchi, "A mutual distortion and impairment compensator for wideband direct-conversion transmitters using neural networks," *IEEE Transactions on Broadcasting*, vol. 58, pp. 168-177, 2012.
- [125] M. Rawat, K. Rawat, M. Younes, and F. Ghannouchi, "Joint mitigation of nonlinearity and modulator imperfections in dual-band concurrent transmitter using neural networks," *Electronics letters*, vol. 49, pp. 253-255, 2013.
- [126] M. Rawat and F. Ghannouchi, "Adaptive linearization of transmitter in the presence of I/Q Imbalance using distributed spatio-temporal neural network," *IEEE International Conference on Wireless Information Technology and Systems*, pp. 1-4, 2012.
- [127] S. Boumaiza and F. M. Ghannouchi, "Realistic power-amplifiers characterization with application to baseband digital predistortion for 3G base stations," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 3016-3021, 2002.
- [128] S. Boumaiza, M. Helaoui, O. Hammi, T. Liu, and F. M. Ghannouchi, "Systematic and adaptive characterization approach for behavior modeling and correction of dynamic nonlinear transmitters," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, pp. 2203-2211, 2007.
- [129] D. Wisell, D. Ronnow, and P. Handel, "A technique to extend the bandwidth of an RF power amplifier test bed," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, pp. 1488-1494, 2007.
- [130] *Fundamentals of Real-Time Spectrum Analysis*: Tektronix Incorporation, 2005.
- [131] D. Huang, H. Leung, and X. Huang, "Experimental evaluation of predistortion techniques for high-power amplifier," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, pp. 2155-2164, 2006.
- [132] Y. Rolain, W. Van Moer, J. Schoukens, and T. Dhaene, "Estimation and validation of semiparametric dynamic nonlinear models," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, pp. 395-400, 2008.
- [133] C. Fager and K. Andersson, "Improvement of oscilloscope based RF measurements by statistical averaging techniques," *IEEE MTT-S International Microwave Symposium Digest*, pp. 1460-1463, 2006.
- [134] F. Macraigne, T. Reveyrand, C. Maziere, D. Barataud, J. Nebus, R. Quere, *et al.*, "A fully calibrated four channels time domain RF envelope measurement system for the envelope characterization of nonlinear devices in a load-pull environment," *European Microwave Conference*, vol. 2, p. 4 pp., 2005.
- [135] L. Angrisani and A. Napolitano, "Modulation quality measurement in WiMAX systems through a fully digital signal processing approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, pp. 2286-2302, 2010.
- [136] K. Ho, Y. Chan, and R. Inkol, "A digital quadrature demodulation system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, pp. 1218-1227, 1996.
- [137] J. G. Proakis and D. G. Manolakis, *Digital signal processing: principles, algorithms, and applications* vol. 3: Prentice Hall New Jersey, 1996.

- [138] B. D. Laki and C. J. Kikkert, "Adaptive Digital Predistortion for Wideband High Crest Factor Applications Based on the WACP Optimization Objective: An Extended Analysis," *IEEE Transactions on Broadcasting*, vol. 59, pp. 136-145, 2013.
- [139] L. Hanzo and T. Keller, *OFDM and MC-CDMA: a primer*: Wiley Online Library, 2006.
- [140] M. O. Pun, M. Morelli, and C. C. J. Kuo, *Multi-carrier techniques for broadband wireless communications: a signal processing perspective*: Imperial College Pr, 2007.
- [141] T. Jiang, L. Song, and Y. Zhang, "Orthogonal Frequency Division Multiple Access Fundamentals and Applications," *Auerbach Publications Boston, MA, USA*, p. 660, 2010.
- [142] M. Lewis and D. M. Theobald, "Preamble aided synchronization," US Patent 7 379 519 Patent, 2008.
- [143] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Communications Letters*, vol. 3, pp. 75-77, 1999.
- [144] M. Iwasaki, "Carrier regenerating device correctly operable in mobile satellite communication," US Patent 5 363 415 Patent, 1994.
- [145] J. Lei, D. Gang, and Z. Ping, "EVM measurement algorithm for OFDM transmitters," *International Symposium on Communications and Information Technologies*, pp. 102-107, 2006.
- [146] A. Schuster, "On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena," *Terrestrial Magnetism*, vol. 3, pp. 13-41, 1898.
- [147] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, pp. 70-73, 1967.
- [148] M. Ghaderi and S. Kumar, "Power Amplifier Adaptive Linearization Using Predistortion with Polynomial Functions," *22nd European Microwave Conference*, vol. 2, pp. 1125-1130, 1992.
- [149] M. Jin, S. K. Shin, D. Oh, and J. Kim, "Reduced order RLS polynomial predistortion," *Proceedings of the 2003 International Symposium on Circuits and Systems*, vol. 4, pp. IV-333-IV-336, 2003.
- [150] T. S. Nielsen, S. Tawfik, T. Larsen, and S. Lindfors, "Behavioral level modeling of power amplifiers with varying antenna load," *IEEE 61st Vehicular Technology Conference*, vol. 2, pp. 963-967, 2005.
- [151] M. O'Droma, S. Meza, and Y. Lei, "New modified Saleh models for memoryless nonlinear power amplifier behavioural modelling," *IEEE Communications Letters*, vol. 13, pp. 399-401, 2009.
- [152] F. Mkadem and S. Boumaiza, "Extended Hammerstein behavioral model using artificial neural networks," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, pp. 745-751, 2009.
- [153] T. M. Ylamurto, "Frequency domain IQ imbalance correction scheme for orthogonal frequency division multiplexing (OFDM) systems," *IEEE Wireless Communications and Networking Conference*, vol. 1, pp. 20-25, 2003.
- [154] N. Sivannarayana and K. Veerabhadra Rao, "IQ imbalance correction in time and frequency domains with application to pulse doppler radar," *Sadhana*, vol. 23, pp. 93-102, 1998.
- [155] K. Sun, I. Darwazeh, L.-K. Huang, and A. Jones, "Optimal pilot based frequency-dependent I/Q imbalance compensation for wideband direct-conversion transmitters," *IEEE Wireless Communications and Networking Conference*, pp. 226-231, 2012.
- [156] F. J. Doyle, R. K. Pearson, and B. A. Ogunnaike, *Identification and control using Volterra models*: Springer Verlag, 2002.
- [157] T. H. Michael, *Scientific Computing: An Introductory Survey*: McGraw-Hill, 2002.
- [158] G. Li and G. L. Stèuber, *Orthogonal frequency division multiplexing for wireless communications*: Springer, 2006.
- [159] M. M. Lajos L. Hanzo, B. J. Choi, Thomas Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs, and broadcasting*: Wiley, 2003.

- [160] J. R. Feigin, "Methods of I/Q imbalance correction for packet-switched IEEE802. 11-compliant OFDM radios," Electrical Engineering Dissertations, Northeastern University, 2008.
- [161] T.-D. Chiueh, P.-Y. Tsai, and I.-W. Lai, *Baseband receiver design for wireless MIMO-OFDM communications*: Wiley, 2012.
- [162] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, pp. 14-37, 1992.
- [163] H. Sari, G. Karam, and I. Jeanclaude, "Frequency-domain equalization of mobile radio and terrestrial broadcast channels," *IEEE Global Telecommunications Conference*, pp. 1-5, 1994.
- [164] A. Czylik, "Comparison between adaptive OFDM and single carrier modulation with frequency domain equalization," *IEEE 47th Vehicular Technology Conference*, vol. 2, pp. 865-869, 1997.
- [165] N. Al-Dhahir, "Single-carrier frequency-domain equalization for space-time block-coded transmissions over frequency-selective fading channels," *IEEE Communications Letters*, vol. 5, pp. 304-306, 2001.
- [166] M. K. Tsatsanis, "Inverse filtering criteria for CDMA systems," *IEEE Transactions on Signal Processing*, vol. 45, pp. 102-112, 1997.
- [167] L. Hanzo, M. Münster, B.-J. Choi, and T. Keller, *OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting*: John Wiley & Sons, 2003.
- [168] T.-D. Chiueh and P.-Y. Tsai, *OFDM baseband receiver design for wireless communications*: John Wiley & Sons, 2008.
- [169] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*: Society for Industrial Mathematics, 1998.
- [170] R. D. DeGroat and E. M. Dowling, "Data least squares problem and channel equalization," *IEEE Transactions on signal processing*, vol. 41, pp. 407-411, 1993.
- [171] A. J. Van Der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1136-1155, 1996.
- [172] G. Cirrincione, G. Ganesan, K. Hari, and S. Van Huffel, "Direct and neural techniques for the data least squares problem," in *International Symposium on the Mathematical Theory of Networks and Systems*, Perpignan, France, 2000.
- [173] J. S. Lim, "A Neural Network Solution on Data Least Square Algorithm and Its Application for Channel Equalization," *Advances in Neural Networks-ISNN*, pp. 678-685, 2007.
- [174] J. S. Lim, "Recursive DLS solution for extreme learning machine-based channel equalizer," *Neurocomputing*, vol. 71, pp. 592-599, 2008.
- [175] X. W. Chang, G. H. Golub, and C. C. Paige, "Towards a backward perturbation analysis for data least squares problems," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, pp. 1281-1301, 2009.
- [176] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, pp. 211-218, 1936.
- [177] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Mathematik*, vol. 14, pp. 403-420, 1970.
- [178] K. K. Htike and O. O. Khalifa, "Rainfall forecasting models using focused time-delay neural networks," *International Conference on Computer and Communication Engineering*, pp. 1-6, 2010.
- [179] T. Clarke, "Generalization of neural networks to the complex plane," *International Joint Conference on Neural Networks*, pp. 435-440, 1990.
- [180] M. S. Kim and C. C. Guest, "Modification of backpropagation networks for complex-valued signal processing in frequency domain," *International Joint Conference on Neural Networks*, pp. 27-31, 1990.

- [181] H. Leung and S. Haykin, "The complex backpropagation algorithm," *IEEE Transactions on Signal Processing*, vol. 39, pp. 2101-2104, 1991.
- [182] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," *Parallel Distributed Processing*, vol. 1, pp. 318-362, 1986.
- [183] M. Arai, "Mapping abilities of three-layer neural networks," *International Joint Conference on Neural Networks*, pp. 419-423, 1989.
- [184] O. Hammi, S. Boumaiza, and F. M. Ghannouchi, "On the robustness of digital predistortion function synthesis and average power tracking for highly nonlinear power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, pp. 1382-1389, 2007.
- [185] M. Bruno, F. Gregorio, J. Cousseau, A. Ghadam, and M. Valkama, "A novel predistorter for highly nonlinear broadband power amplifier," *Argentine School of Micro-Nanoelectronics, Technology and Applications* pp. 84-89, 2012.
- [186] W. Bridewell, N. B. Asadi, P. Langley, and L. Todorovski, "Reducing overfitting in process model induction," *ACM Proceedings of the 22nd international conference on Machine learning*, pp. 81-88, 2005.
- [187] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, pp. 303-314, 1989.
- [188] K.-I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural networks*, vol. 2, pp. 183-192, 1989.
- [189] S. Van Huffel and J. Vandewalle, *The total least squares problem: computational aspects and analysis* vol. 9: Society for Industrial Mathematics, 1991.
- [190] P. P. Gallo, "Consistency of regression estimates when some variables are subject to error," *Communications in Statistics-Theory and Methods*, vol. 11, pp. 973-983, 1982.
- [191] L. J. Gleser, "Estimation in a multivariate" errors in variables" regression model: large sample results," *The Annals of Statistics*, pp. 24-44, 1981.
- [192] C. C. Paige and Z. Strakoš, "Unifying least squares, total least squares and data least squares," *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*, p. 25, 2002.
- [193] G. H. Golub and C. F. Van Loan, *Matrix computations*: Johns Hopkins Univ Press, 1996.
- [194] E. H. Bareiss, "Numerical solution of linear equations with Toeplitz and vector Toeplitz matrices," *Numerische Mathematik*, vol. 13, pp. 404-424, 1969.
- [195] W. H. Press, *Numerical recipes: The art of scientific computing*: Cambridge university press, 2007.
- [196] P. J. Davis, *Circulant matrices*: John Wiley & Sons, 1979.

Appendix A Least Squares

A.1 Traditional LS problem

S.V. Huffel et al. [189] gave a definition for least squares problem as:

Definition A-1 For a $N \times L$ matrix $\mathbf{A}$ and the corresponding overdetermined linear equation $\mathbf{Ax} \simeq \mathbf{b}$, its least squares solution is to find

$$\min_{\mathbf{b}' \in \mathbb{C}^N} \|\mathbf{b} - \mathbf{b}'\|_2 \quad (\text{A.1})$$

subject to $\mathbf{b}' \in \text{span}(\mathbf{A})$

And once $\mathbf{b}'$ is found, the $\mathbf{x}$ satisfying $\mathbf{Ax} = \mathbf{b}'$ is the least squares solution. $\square$

A.2 Total Least Squares

The limitation of traditional least squares method is that it implicitly assumes that the error occurs only in vector $\mathbf{b}$ of equation $\mathbf{Ax} \simeq \mathbf{b}$. The limitation can be seen in .

The assumption cannot reflect the real conditions in many cases, esp. in the case of predistortion of power amplifier or IQ imbalance. To model the inverse function of power amplifier or IQ imbalance, the output and the input are often interchanged, resulting in noisy matrix $\mathbf{A}$ of equation $\mathbf{Ax} \simeq \mathbf{b}$. Or the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ can be both polluted by noise, when both the input and the output signals are obtained from real noisy signals. Hence a method to handle noisy $\mathbf{A}$ is necessary.

Total least squares (TLS) problem is to find an optimal solution of equation $\mathbf{Ax} \simeq \mathbf{b}$ under the assumption that both the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are perturbed by noise.

Definition A-2 For an overdetermined equation $\mathbf{Ax} \simeq \mathbf{b}$, its total least squares solution is to find

$$\min_{[\mathbf{A}'; \mathbf{b}'] \in \mathbb{C}^{N \times (L+1)}} \|[\mathbf{A}; \mathbf{b}] - [\mathbf{A}'; \mathbf{b}']\|_F \quad (\text{A.2})$$

subject to $\mathbf{b}' \in \text{span}(\mathbf{A}')$. And for any $[\mathbf{A}'; \mathbf{b}']$ minimizing (A.2), the solution $\mathbf{x}$ of

$$\mathbf{A}'\mathbf{x} = \mathbf{b}' \quad (\text{A.3})$$

is the TLS solution of the original equation $\mathbf{Ax} \simeq \mathbf{b}$. Please note in (A.2), the function $\|\cdot\|_F$ is *Frobenius* norm defined in (6.4). $\square$

A fundamental property of TLS solution is its consistency with zero mean, independently and identically distribution (IID) perturbation. Assume $[\mathbf{A}_0; \mathbf{b}_0]$ is perturbed by $[\Delta\mathbf{A}; \Delta\mathbf{b}]$. Moreover, the perturbation $[\Delta\mathbf{A}; \Delta\mathbf{b}]$ is IID, with zero mean and covariance matrix $\eta\mathbf{I}$. Then according to [190] and [191], TLS solution is consistent, i.e., the TLS solution converges to the noise-free solution $(\mathbf{A}_0^H \mathbf{A}_0)^{-1} \mathbf{A}_0^H \mathbf{b}_0$ as the number of samples $N \rightarrow \infty$.

A.3 Data Least Squares

The ordinary least squares problems assume that the right side $\mathbf{b}$ of equation $\mathbf{Ax} \simeq \mathbf{b}$ suffers noise while the left side does not. The total least squares problems assume that both sides of equation $\mathbf{Ax} \simeq \mathbf{b}$ suffer noise. Now the third type of least squares problems – data least squares (DLS) problems – are based on the assumption that only the left side of equation $\mathbf{Ax} \simeq \mathbf{b}$ suffers noise while the right side is noise free. R. D. DeGroat et al. [170] describe DLS problems as:

$$(\mathbf{A} + \mathbf{E})\mathbf{x} \simeq \mathbf{b} \quad (\text{A.4})$$

where $\mathbf{E}$ is the noisy perturbation. Then G. Cirrincione et al. [172] define DLS problems as:

Definition A-3 For an overdetermined equation $\mathbf{Ax} \simeq \mathbf{b}$, its data least squares solution is to find

$$\min_{\mathbf{A}' \in \mathbb{C}^{N \times L}} \|\mathbf{A} - \mathbf{A}'\|_F \quad (\text{A.5})$$

subject to $\mathbf{b} \in \text{span}(\mathbf{A}')$

And for any $\mathbf{A}'$ minimizing (A.5), the solution $\mathbf{x}$ of

$$\mathbf{A}'\mathbf{x} = \mathbf{b} \quad (\text{A.6})$$

is the DLS solution of the original equation $\mathbf{Ax} \simeq \mathbf{b}$. $\square$

Then there is a way to unify the three kinds of least squares problems under one definition [172, 192].

Definition A-4 For an overdetermined equation $\mathbf{Ax} \simeq \mathbf{b}$, the scaled total least squares (STLS) solution is to find

$$\min_{[\Delta\mathbf{A}; \Delta\mathbf{b}] \in \mathbb{C}^{N \times (L+1)}} \left\| [\Delta\mathbf{A}; \mu\Delta\mathbf{b}] \right\|_F \quad (\text{A.7})$$

subject to $\mathbf{b} + \Delta\mathbf{b}' \in \text{span}(\mathbf{A} + \Delta\mathbf{A}')$

where $\mu > 0$. And for any $[\Delta\mathbf{A}; \Delta\mathbf{b}]$ minimizing (A.7), the solution $\mathbf{x}$ of

$$(\mathbf{A} + \Delta\mathbf{A})\mathbf{x} = (\mathbf{b} + \Delta\mathbf{b}) \quad (\text{A.8})$$

is the STLS solution of the original equation $\mathbf{Ax} \simeq \mathbf{b}$. $\square$

Theorem 7-1 Depending on the value of μ , the STLS solution is shown to be

If $\mu \rightarrow 0$, $\mathbf{x}$ is the ordinary least squares solution;

If $\mu = 1$, $\mathbf{x}$ is the total least squares solution;

If $\mu \rightarrow +\infty$, $\mathbf{x}$ is the data least squares solution;

Proof. For the proof see [192]. $\square$

Data least squares problems and their solutions are very useful in identifying the model coefficients on predistortion of power amplifier or IQ imbalance. Due to the nature of predistortion – the inverse relationship of DUT – the real input and output signals of DUT are interchanged in the process of identification. Hence there are cases that the left side matrix $\mathbf{A}$ of equation $\mathbf{Ax} \simeq \mathbf{b}$ suffers noise while the right side vector $\mathbf{b}$ is noise-free. The identification is a DLS problem.

Appendix B Singular Value Decomposition

Some fundamental properties of SVD are discussed here, including its own sensitivity to noise or perturbation.

Theorem 7-2 An overdetermined $N \times L$ matrix $\mathbf{A}$, either full column rank or not, can be decomposed in the following form:

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}^H \quad (\text{B.1})$$

where $\mathbf{U}$ is an $N \times N$ real or complex unitary matrix, $\mathbf{V}$ is an $L \times L$ real or complex unitary matrix, and Σ is $N \times L$ diagonal matrix with its diagonal elements being all nonnegative real numbers.

Proof. For the proof see [193]. $\square$

The diagonal element σ_i of Σ is called *singular value*, the column vector $\mathbf{u}_i$ of $\mathbf{U}$ and the column vector $\mathbf{v}_i$ of $\mathbf{V}$ are called the *left singular vector* and *right singular vector* respectively. And for nonzero singular value σ_i

$$\begin{cases} \mathbf{A} \mathbf{v}_i = \sigma_i \mathbf{u}_i \\ \mathbf{A}^H \mathbf{u}_i = \sigma_i \mathbf{v}_i \end{cases} \quad (\text{B.2})$$

Moreover, $\mathbf{u}_i$ are the eigenvector of $\mathbf{A} \mathbf{A}^H$, corresponding to σ_i^2 . And $\mathbf{v}_i$ are the eigenvector of $\mathbf{A}^H \mathbf{A}$, corresponding to σ_i^2 .

The singular values of a $N \times L$ matrix $\mathbf{A}$ can be reordered so that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_L \geq 0$. And if $\mathbf{A}$ is full column rank, all singular values are positive.

Otherwise, if $\text{rank}(\mathbf{A}) = p < L$, then $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p > \sigma_{p+1} = \dots = \sigma_L = 0$. Moreover, the singular vectors provide a good representation of the span space and null space of matrix $\mathbf{A}$

$$\begin{cases} \text{span}(\mathbf{A}) = \text{span}([\mathbf{u}_1, \dots, \mathbf{u}_p]) \\ \text{null}(\mathbf{A}) = \text{span}([\mathbf{v}_{p+1}, \dots, \mathbf{v}_L]) \\ \text{span}(\mathbf{A}^T) = \text{span}([\mathbf{v}_1, \dots, \mathbf{v}_p]) \\ \text{null}(\mathbf{A}^T) = \text{span}([\mathbf{u}_{p+1}, \dots, \mathbf{u}_L]) \end{cases} \quad (\text{B.3})$$

Another important property of SVD is its ability to decompose a matrix into the sum of several 1-rank matrices. It is easy to show that

$$\begin{aligned} \mathbf{A} &= \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^H + \dots + \sigma_p \mathbf{u}_p \mathbf{v}_p^H \\ &= \sigma_1 \mathbf{E}_1 + \sigma_2 \mathbf{E}_2 + \dots + \sigma_p \mathbf{E}_p \end{aligned} \quad (\text{B.4})$$

where $\mathbf{E}_i = \mathbf{u}_i \mathbf{v}_i^H$ is $N \times L$ matrix of rank 1.

Then there is a way to approximate a matrix $\mathbf{A}$ of rank p by another matrix $\tilde{\mathbf{A}}$ of rank $k < p$ with minimum error. According to C. Eckart and G. Young[176], the largest k singular values and the corresponding singular vectors can be used to approximate the matrix of higher rank by a matrix of lower rank at the minimum cost among all matrices of the same rank.

Theorem 7-3 An $N \times L$ matrix $\mathbf{A}$ of rank p can be decomposed into

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^H + \dots + \sigma_p \mathbf{u}_p \mathbf{v}_p^H$$

And an approximation matrix $\tilde{\mathbf{A}}$ has the form

$$\tilde{\mathbf{A}} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^H + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^H + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^H$$

where $k < p$. Then for any $N \times L$ matrix $\mathbf{M}$ of rank k , $\tilde{\mathbf{A}}$ is the best one to approximate $\mathbf{A}$ in the sense of minimum *Frobenius* norm

$$\begin{aligned} \min_{\text{rank}(\mathbf{M})=k} \|\mathbf{A} - \mathbf{M}\|_F &= \|\mathbf{A} - \tilde{\mathbf{A}}\|_F \\ &= \sqrt{\sum_{i=k+1}^p \sigma_i^2} \end{aligned} \quad (\text{B.5})$$

Proof. For the proof see [176]. $\square$

Theorem 7-3 justifies the way to disregard the least singular value or set it to zero. Moreover, in the influence of noise, matrix $\mathbf{A}$ would usually appear to be full column rank. However, some trivial singular values may be purely generated by noise; they are the redundancy of the identified model. In this case, disregarding the trivial singular values and setting them to zero can help reduce the influence of noise while approximate the original matrix to a certain level of accuracy.

The sensitivity of SVD to the perturbation in $\mathbf{A}$ directly influences the robustness of the TLS SVD algorithm. Let the SVD of the noise free matrix $\mathbf{C}_0 = [\mathbf{A}_0; \mathbf{b}_0]$ be

$$\mathbf{C}_0 = \mathbf{U}_0 \Sigma_0 \mathbf{V}_0^H \quad (\text{B.6})$$

and the SVD of the noisy matrix $\mathbf{C} = [\mathbf{A}_0 + \Delta \mathbf{A}; \mathbf{b}_0 + \Delta \mathbf{b}]$ be

$$\mathbf{C} = \mathbf{U} \Sigma \mathbf{V}^H \quad (\text{B.7})$$

Then it is easy to shown that

$$\mathbf{C} = \mathbf{U}_0 \mathbf{R}_1 (\Sigma_0 + \Delta \Sigma) \mathbf{R}_2^H \mathbf{V}_0^H \quad (\text{B.8})$$

where the singular values suffer additive perturbation, while the singular vectors suffer rotations, which are

$$\begin{cases} \mathbf{R}_1 = \mathbf{U}_0^H \mathbf{U} \\ \mathbf{R}_2 = \mathbf{V}_0^H \mathbf{V} \end{cases} \quad (\text{B.9})$$

The additive perturbation in singular values are shown in [189] to be bounded by the perturbation in $\mathbf{C}$

$$\begin{cases} \|\Delta \Sigma\|_F \leq \|\Delta \mathbf{C}\|_F \\ \|\Delta \sigma_i\|_2 \leq \|\Delta \mathbf{C}\|_2 \end{cases} \quad (\text{B.10})$$

Hence the singular values are *well-conditioned*.

On the other hand, the rotations on the singular vectors are more complicated, and directly set a boundary on the sensitivity of the TLS solution. S. V. Huffel et al.[189] gave a boundary for TLS solutions:

Theorem 7-4 Let σ_{L+1}^0 and σ_{L+1} be the minimum singular values of the full column rank noise-free matrix $[\mathbf{A}_0; \mathbf{b}_0]$ and the full column rank noisy matrix $[\mathbf{A}_0 + \Delta \mathbf{A}; \mathbf{b}_0 + \Delta \mathbf{b}]$ respectively, and $\mathbf{v}_{L+1}^0$ and $\mathbf{v}_{L+1}$ be the corresponding right singular vectors. And let $\mathbf{x}_0$ and $\mathbf{x}$ be the TLS solution of $\mathbf{A}_0 \mathbf{x}_0 \simeq \mathbf{b}_0$ and $(\mathbf{A}_0 + \Delta \mathbf{A}) \mathbf{x} \simeq \mathbf{b}_0 + \Delta \mathbf{b}$ respectively. If $\mathbf{x}_0^T \mathbf{x} > 0$, then the angular between $\mathbf{x}_0$ and $\mathbf{x}$ is

$$\cos \angle(\mathbf{x}_0, \mathbf{x}) = \frac{\sqrt{(\|\mathbf{x}_0\|_2^2 + 1)(\|\mathbf{x}\|_2^2 + 1)} \cos \angle(\mathbf{v}_{L+1}^0, \mathbf{v}_{L+1}) - 1}{\|\mathbf{x}_0\|_2 \|\mathbf{x}\|_2} \quad (\text{B.11})$$

Proof. For the proof see [189]. Unfortunately, the explicit formula to calculate the boundary on $\cos \angle(\mathbf{v}_{L+1}^0, \mathbf{v}_{L+1})$ has not been found yet.

Appendix C Toeplitz and Circulant Matrices

An $(N+1) \times (N+1)$ toeplitz matrix is define in the following form [193-195]:

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_N \\ h_{-1} & h_0 & h_1 & \cdots & h_{N-1} \\ h_{-2} & h_{-1} & h_0 & \cdots & h_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{-N} & h_{-N+1} & h_{-N+2} & \cdots & h_0 \end{bmatrix} \quad (\text{C.1})$$

where the diagonal is constant and $\mathbf{H}_{i,j} = \mathbf{H}_{i+1,j+1}$. From (C.1), it is obvious that a toeplitz matrix is uniquely determined by its first row vector $[h_0 \ h_1 \ h_2 \ \cdots \ h_N]$ and its first column vector $[h_0 \ h_{-1} \ h_{-2} \ \cdots \ h_{-N}]^T$.

The circulant matrix is a special toeplitz that each row is only circle-shifting one element of its predecessor to the right. An $(N+1) \times (N+1)$ circulant matrix is defined in the following way [193, 196]:

$$\mathbf{H} = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots & h_N \\ h_N & h_0 & h_1 & \cdots & h_{N-1} \\ h_{N-1} & h_N & h_0 & \cdots & h_{N-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_1 & h_2 & h_3 & \cdots & h_0 \end{bmatrix} \quad (\text{C.2})$$

From (C.2), it is obvious that a circulant matrix is uniquely determined by its first column vector $[h_0 \ h_1 \ h_2 \ \cdots \ h_N]$. Moreover, a circulant matrix can be factorized by fast Fourier Transform, as shown in Theorem 7-5.

Theorem 7-5 A $N \times N$ circulant matrix $\mathbf{H}$ can be factorized in the following way [193, 196]:

$$\mathbf{H}_{N \times N} = \mathbf{F}_N^* \text{diag}\{\mathbf{F}_N \mathbf{h}_1\} \mathbf{F}_N \quad (\text{C.3})$$

where $\mathbf{h}_1$ is the first column vector of $\mathbf{H}_{N \times N}$, and $\mathbf{F}_N$ is unitary FFT matrix of size N

$$[\mathbf{F}]_{nk} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi nk}{N}} \quad (\text{C.4})$$

Proof. For the proof see [196]. $\square$