Contributions to Digital Predistortion of Radio-Frequency Power Amplifiers for Wireless Applications

Ansätze zu digitaler Vorverzerrung von Funkfrequenz-Leistungsverstärkern für drahtlose Anwendungen

Der Technischen Fakultät der Universität Erlangen-Nürnberg zur Erlangung des Grades

DOKTOR-INGENIEUR

vorgelegt von Dipl.-Ing. Xin YU

Erlangen-2012
Als Dissertation genehmigt von
der Technischen Fakultät der
Universität Erlangen-Nürnberg

Tag der Einreichung: 09.12.2011
Dekan: Prof. Dr.-Ing. Marion Merklein
Berichterstatter: Prof. Dr.-Ing. Georg Fischer
                        Prof. Dr.-Ing. Jan-Erik Müller
ACKNOWLEDGMENTS

During my PhD studies I feel very lucky to meet so many nice and talented people, who have supported me with great patience and made this work possible at all. I owe my deepest gratitude to each and every one of them for their guidance, friendship, encouragement and help.

First, I am heartily thankful to my advisor, Prof. Dr.-Ing. Georg Fischer, for providing me the opportunity of doing my doctorate at Friedrich-Alexander University. During my doctoral studies he has always been there to give me useful advices and valuable commenting on my views as well as enrich my ideas.

Secondly, I would like to thank my supervisor Dr. Andreas Pascht, who has continuously provided the freedom and support in the research fields as well as nice work environment at Bell-Labs, Alcatel-Lucent. Therefore, I am able to finish my doctorate during my work as research engineer for Alcatel-Lucent.

I also benefit a lot from my outstanding colleagues, Thomas Bohn, Dr. Christoph Haslach, Dirk Wiegner, Robin Machinal, Gerhard Luz, Patrick Jueschke, Simone Maier and Dr. Hong Jiang. Especially Mr. Bohn gave me a lot of practice advices and implementation support in hardware experiments with respect to verifying the novel ideas in this work. My special thanks also go to Dr. Andreas Frotzscher for carefully reading and useful commenting on revisions of my manuscript.

Finally, I am grateful to my family, my parents, and especially my wife, for their understanding, encouragement and financial support. Because of their unconditional love and sacrifices over these years I am able to highly concentrate on my PhD studies.
Abstract

Our daily lives would look completely different without wireless communication. The mobility of access to wireless networks anytime and anywhere has a direct impact on our daily lives nearly at every corner of the world. However, the demand for fast and robust wireless networks increases permanently. One of the key components both for user end-device and base stations is the power amplifier, which is used to amplify the transmit signal, so that the transmit signal can reach the user over tens of kilometres. Unfortunately, power amplifiers suffer from nonlinearities, if they are driven at an effective operating point. This not only results in high bit error rate, but also strong interference for the adjacent communication channels. Because of the specifications for the amount of in-band and out-band distortion of transmit signals, such distorted signals are forbidden to be emitted in the air. If power amplifiers work only in linear region, the specifications defined by the governments or international telecommunication associations would be met. However, in this case the power amplifiers would have very low efficiency.

Nowadays, the subject of climate change experiences a clear and long overdue attention around the world and one therefore cannot disregard the extreme low energy efficiency of the power amplifiers. The solution for efficiency improvement of power amplifiers are linearization techniques. By using these linearization techniques one is able to operate the power amplifiers in a nonlinear region with often high efficiency and satisfies the specifications given by standards at the same time. Digital predistortion is one of the most popular linearization techniques because of its overall efficiency and outstanding performance. Digital predistortion is a kind of digital signal processing and is located in the digital baseband. The nonlinearities caused by power amplifiers can be almost completely compensated by digital predistortion.

In this work the digital predistortion technique is investigated with respect to system stability, computational cost and adaptation flexibility. Novel calculation approaches are developed to simultaneously improve the stability and reduce the complexity of the digital predistortion system. Furthermore, in some applications fast load changes raise a series of difficulties for
digital predistortion. Therefore, the case of fast load changes is also addressed in this work and a new approach is developed to deal with this problem. Moreover, many innovative ideas for efficiency enhancement are introduced into power amplifier design in the last decade, such as Envelope-Tracking. Besides the nonlinearities of power amplifiers the Envelope Tracking system introduces another problem of temporal mismatch between the RF input signal and its dynamically changing power supply voltage. Special analysis and compensation approaches for this temporal mismatch are presented in this work.
Zusammenfassung


TABLE OF CONTENTS

Acknowledgments.................................................................................................................................I
Table of Contents ..................................................................................................................................VI
List of Figures......................................................................................................................................... IX
Glossary .................................................................................................................................................. XII
Chapter 1 Introduction .......................................................................................................................... 1
Chapter 2 Challenges with Digital Predistortion.................................................................................. 4
  2.1 Research Tasks.............................................................................................................................. 4
    2.1.1 PA Memory Delay Analysis .............................................................................................. 4
    2.1.2 Reliability and Stability of Digital Predistortion.............................................................. 5
    2.1.3 Complexity and Computational Cost of Digital Predistortion........................................ 5
    2.1.4 Fast Load Changing of PAs ............................................................................................ 6
    2.1.5 Temporal Mismatch in Envelope-Tracking System........................................................... 7
  2.2 Major Contributions...................................................................................................................... 7
  2.3 Patents and Publications.............................................................................................................. 8
  2.4 Literature Review......................................................................................................................... 10
Chapter 3 Tools for Assessing Performance ......................................................................................... 13
  3.1 Simulation Tools.......................................................................................................................... 13
  3.2 Hardware Test Platform ............................................................................................................. 13
    3.2.1 PowerPC System .............................................................................................................. 15
    3.2.2 Offline Digital Predistortion of Multi-LUT Structure....................................................... 16
Chapter 4 PA Models and Predistortion Algorithms .......................................................................... 18
  4.1 PA Behavioural Models .............................................................................................................. 18
    4.1.1 Memoryless PA Models ..................................................................................................... 19
    4.1.2 Memory PA Models .......................................................................................................... 21
      4.1.2.1 Short Term Memory Effects .................................................................................... 22
      4.1.2.2 Long Term Memory Effects .................................................................................... 27
  4.2 Power Amplifier Linearization Techniques ................................................................................. 28
    4.2.1 Feedback Technique ........................................................................................................... 28
    4.2.2 Feedforward Technique .................................................................................................... 29
4.2.3 Predistortion Techniques ................................................................. 29
4.2.4 Digital Predistortion Technique ...................................................... 31
  4.2.4.1 Memoryless Digital Predistortion ............................................. 35
  4.2.4.2 Memory Digital Predistortion .................................................... 38
  4.2.4.3 Online Digital Predistortion .................................................... 40
  4.2.4.4 Digital Predistortion for Mobile Phone Devices ..................... 42
4.3 Memory Delay Estimation by Using Two-Tone Test ......................... 42
  4.3.1 Theory ......................................................................................... 44
  4.3.2 Simulation Results ........................................................................ 48
  4.3.3 Experimental Results ................................................................. 51
4.4 Coefficient Calculation by Using Stationary Iterative Methods .......... 53
  4.4.1 Stability Estimation of Digital Predistortion System ................... 54
  4.4.2 The First-Order Stationary Iterative Methods ............................... 57
  4.4.3 The Second-Order Stationary Iterative Methods ......................... 61
  4.4.4 Simulation Results ........................................................................ 62
  4.4.5 Experimental Results ................................................................. 64
4.5 Digital Predistortion Using Adaptive Selection of Basis Functions ... 67
  4.5.1 Indirect Learning DPD Using Adaptive Selection of Basis Functions . 70
  4.5.2 Direct Learning DPD Using Adaptive Selection of Basis Functions .. 72
  4.5.3 Simulation Results ........................................................................ 74
  4.5.4 Experimental Results ................................................................. 79
4.6 Combination of Adaptive Selection of Basis Functions and Iterative Methods . 82
4.7 Fast Load Changing of Power Amplifiers ........................................... 82
  4.7.1 Problem Analysis ........................................................................ 83
  4.7.2 Method of Attaching Old Samples .............................................. 84
  4.7.3 Experimental Results ................................................................. 86
4.8 ET Temporal Misalignment Analysis and its Compensation Methods .... 93
  4.8.1 Analysis in Time Domain ............................................................. 95
  4.8.2 Analysis in Frequency Domain .................................................... 98
  4.8.3 Experimental Results ................................................................. 105
Chapter 5 Complexity Assessment .......................................................... 108
  5.1 Computational Cost of Stationary Iterative Methods ....................... 108
  5.2 Computational Cost of DPD Using Adaptive Selection of Basis Functions .... 109
Chapter 6 Conclusion .............................................................................. 111
6.1 Contributions ....................................................................................................................111
6.2 Key Results ........................................................................................................................112
6.3 Suggestions for Future Research....................................................................................113

Appendix A. Memory Delay Estimation for PA with Memory Polynomial Model ............... i
Appendix B. Proof of Positive-Semi-Definite Matrix ............................................................... v
Appendix C. Relationship between Temporal Mismatch and Unsymmetrical IMDs ........... vi
List of References......................................................................................................................... ix
LIST OF FIGURES

Figure 1.1: Power consumption of a GSM base station. ................................................................. 1
Figure 1.2: PA energy efficiency roadmap without using of linearization and reduction of signal
dynamic .......................................................................................................................................... 2
Figure 2.1: Dynamic load change of OFDM signal. ......................................................................... 6
Figure 3.1: Transmitter block diagram ......................................................................................... 14
Figure 3.2: Hardware platform composed of Xilinx evaluation board, adaptation board, analogue
feedforward and analogue feedback boards ..................................................................................... 15
Figure 3.3: Transmitter overview including control system ........................................................... 15
Figure 3.4: Offline DPD of multi-LUT structure ............................................................................. 16
Figure 3.5: Flow chart of offline DPD .............................................................................................. 17
Figure 4.1: Simplified transmitter structure ................................................................................... 19
Figure 4.2: Typical AM-AM and AM-PM characteristics of a Class-AB PA ..................................... 20
Figure 4.3: The Wiener model ......................................................................................................... 23
Figure 4.4: The Hammerstein model .............................................................................................. 23
Figure 4.5: The Wiener-Hammerstein model .................................................................................. 24
Figure 4.6: Memory Polynomial model .......................................................................................... 25
Figure 4.7: Auto-Regressive Moving Average model ..................................................................... 26
Figure 4.8: PA model with long memory effects ............................................................................ 27
Figure 4.9: PA model with long memory effects using feedback loop ........................................... 28
Figure 4.10: Feedback technique .................................................................................................... 29
Figure 4.11: Feedforward technique ............................................................................................... 29
Figure 4.12: Principle of predistortion ............................................................................................. 30
Figure 4.13: Conventional DPD in baseband .................................................................................. 31
Figure 4.14: Hybrid Digital/RF envelope predistortion ................................................................... 31
Figure 4.15: Direct learning and indirect learning architectures of DPD system ............................ 33
Figure 4.16: Difference between online DPD and offline DPD ....................................................... 34
Figure 4.17: Open loop DPD ........................................................................................................... 35
Figure 4.18: Comparison of PA modelling and indirect learning DPD ............................................ 36
Figure 4.19: Indirect learning DPD of memory polynomial structure ............................................. 39
Figure 4.20: Online DPD with LUT structure .................................................................................. 41
Figure 4.21: PA RF memory polynomial model .............................................................................. 44
Figure 4.22: Flow chart of PA memory delay estimation ................................................................. 49
Figure 4.23: Estimated memory delay constant of PA memory polynomial model with 2 taps ....... 50
Figure 4.24: Estimated memory delay constant of PA memory polynomial model with 3 taps ....... 51
Figure 4.25: Condition number of memoryless DPD with 8096 samples ...................................... 55
Figure 4.26: Condition number of memoryless DPD with 8096 samples ...................................... 55
Figure 4.27: DPD performance vs. maximal number of iterations for Gauss Seidel method ........ 62
Figure 4.28: DPD performance vs. maximal residual for Gauss Seidel method ............................ 63
Figure 4.29: PA output spectrum with DPD solved by Gauss-Seidel method. ..........................65
Figure 4.30: Orthogonal polynomial DPD characteristics of LS method and Gauss-Seidel method
for a single-carrier UMTS signal. ..................................................................................66
Figure 4.31: Orthogonal polynomial DPD characteristics of LS method and Gauss-Seidel method
for a 3-carrier UMTS signal. ..................................................................................66
Figure 4.32: PA output spectrum of indirect learning DPD of using adaptive selection of basis
functions. ..................................................................................................................74
Figure 4.33: Number of basis functions used in DPD adaptation. ........................................76
Figure 4.34: Condition number of indirect learning DPD of using adaptive selection of basis
functions. ..................................................................................................................76
Figure 4.35: PA output spectrum of indirect learning DPD with relaxed threshold. ..................77
Figure 4.36: Number of basis functions of indirect learning DPD with relaxed threshold. ........78
Figure 4.37: PA output spectrum of direct learning DPD of using adaptive selection of basis
functions. ..................................................................................................................78
Figure 4.38: Number of basis functions used in DPD adaptation of direct learning architecture........79
Figure 4.39: Condition number of direct learning DPD using adaptive selection of basis functions. ..79
Figure 4.40: PA output spectrum of direct learning DPD of using adaptive selection of basis
functions. ..................................................................................................................80
Figure 4.41: Number of basis functions used in DPD adaptation of direct learning architecture.......81
Figure 4.42: Condition number of direct learning DPD of using adaptive selection of basis
functions. ..................................................................................................................81
Figure 4.43: HS-DSCH code and time structures.................................................................83
Figure 4.44: Multi-DPD approach for dealing with fast load change. .....................................85
Figure 4.45: Method of attaching old samples of missing power level to new captured samples.......85
Figure 4.46: Flow chart of attaching old samples. .................................................................87
Figure 4.47: I/Q test signal with 2ms Transmission Time Interval of full load and 25% load. ........87
Figure 4.48: Amplitude of captured samples of full load and 25% load. .................................88
Figure 4.49: PA output spectra with signal burst of difference load cases (1). .........................88
Figure 4.50: PA output spectra with signal burst of difference load cases (2). .........................89
Figure 4.51: Amplitude of captured samples of 25% load plus old samples. ............................90
Figure 4.52: PA output spectra with signal burst of difference load cases using method of attaching
old samples. .............................................................................................................90
Figure 4.53: PA output spectra using method of attaching old samples and weighting factor for
attached samples. .....................................................................................................91
Figure 4.54: Estimated DPD characteristics with and without weighting factor..........................92
Figure 4.55: PA output spectra without attaching old samples. .............................................92
Figure 4.56: PA output spectra with attaching of old samples................................................93
Figure 4.57: ET system. ...................................................................................................94
Figure 4.58: Symmetrical signal amplified by ET system. ....................................................96
Figure 4.59: Compensation of temporal mismatch with integer delay......................................97
Figure 4.60: Compensation of temporal mismatch with fractional delay.................................98
Figure 4.61: AM-PM distortion of a PA based on a GaN transistor. .........................................99
Figure 4.62: Novel PA behaviour model for ET system. .......................................................100
Figure 4.63: AM-PM distortion of a parabolic form...............................................................101
Figure 4.64: Spectrum V.S. temporal mismatch between RF signal and its ET signal..................105
Figure 4.65: ACP Asymmetry over temporal misalignment between RF signal and its ET signal...105
Figure 4.66: Measured ACP asymmetry vs. temporal mismatch.............................................106
Figure 4.67: Flow chart of signal conditioning for ET system..................................................107
Figure 4.68: Spectrum improvement after temporal mismatch compensation..........................107
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2G</td>
<td>Second-Generation</td>
</tr>
<tr>
<td>3G</td>
<td>Third-Generation</td>
</tr>
<tr>
<td>4G</td>
<td>Fourth-Generation</td>
</tr>
<tr>
<td>ACLR</td>
<td>Adjacent Channel Leakage Power Ratio</td>
</tr>
<tr>
<td>ACP</td>
<td>Adjacent Channel Power</td>
</tr>
<tr>
<td>ADC</td>
<td>Analogue Digital Converter</td>
</tr>
<tr>
<td>AM-AM</td>
<td>Amplitude-Dependent Amplitude Distortion</td>
</tr>
<tr>
<td>AM-PM</td>
<td>Amplitude-Dependent Phase Distortion</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>CW</td>
<td>Continuous Wave</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital Analogue Converter</td>
</tr>
<tr>
<td>DPD</td>
<td>Digital Predistortion</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processor</td>
</tr>
<tr>
<td>DUT</td>
<td>Device Under Test</td>
</tr>
<tr>
<td>ET</td>
<td>Envelope Tracking</td>
</tr>
<tr>
<td>FPGA</td>
<td>Field Programmable Gate Array</td>
</tr>
<tr>
<td>GaN</td>
<td>Gallium Nitride</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphical User Interface</td>
</tr>
<tr>
<td>HFET</td>
<td>Heterostructure Field Effect Transistor</td>
</tr>
<tr>
<td>HSDPA</td>
<td>High Speed Downlink Packet Access</td>
</tr>
<tr>
<td>HS-SDCH</td>
<td>High Speed Downlink Shared Channel</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IMD</td>
<td>Intermodulation Distortion</td>
</tr>
<tr>
<td>LS</td>
<td>Least Square Method</td>
</tr>
<tr>
<td>LMS</td>
<td>Least Mean Square Method</td>
</tr>
<tr>
<td>LO</td>
<td>Local Oscillator</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>LTI</td>
<td>Linear Time-Invariant</td>
</tr>
<tr>
<td>MB</td>
<td>Megabytes</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency-Division Multiplexing</td>
</tr>
<tr>
<td>PA</td>
<td>Power Amplifier</td>
</tr>
<tr>
<td>PAPR</td>
<td>Peak-to-Average Power Ratio</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>PD</td>
<td>Predistorter</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>RBW</td>
<td>Resolution Bandwidth</td>
</tr>
<tr>
<td>RF</td>
<td>Radio Frequency</td>
</tr>
<tr>
<td>RIO</td>
<td>Ratio of In-band and Out of band emission power</td>
</tr>
<tr>
<td>SoC</td>
<td>System on a Chip</td>
</tr>
<tr>
<td>SDRAM</td>
<td>Synchronous Dynamic Random Access Memory</td>
</tr>
<tr>
<td>SPI</td>
<td>Serial Peripheral Interface</td>
</tr>
<tr>
<td>TTI</td>
<td>Transmission Time Interval</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>VHDL</td>
<td>Very High Speed Integrated Circuit Hardware Description Language</td>
</tr>
<tr>
<td>VBW</td>
<td>Video Bandwidth</td>
</tr>
<tr>
<td>WCDMA</td>
<td>Wideband Code Division Multiple Access</td>
</tr>
</tbody>
</table>
Chapter 1 Introduction

In wireless communication systems the radio frequency (RF) power amplifiers (PAs) are very critical components in terms of power consumption. Without using any linearization techniques the PAs have very low efficiency and consume the most energy in a base station. The power consumption of a typical GSM base station is depicted in Fig. 1.1. The total energy consumption of this base station amounts to 1780W and only 540W useful RF power is emitted in the air. Thereby, the largest energy loss about 900W in this base station is caused by the PA, which has only 37% power efficiency.

Since the transmit GSM signal has constant envelope, the PA can be operated in nonlinear region with relatively high power efficiency. But for the signal with dynamical envelope the PA must therefore be operated in linear region for the entire signal power range. Thus the efficiency of these PAs would be much worse in comparison to the PA efficiency of a typical GSM base station. In most cases the power efficiency of PAs in these applications with dynamical envelope would be lower than 20% without using of any linearization techniques. Moreover, because the use of smart-phones is an ongoing trend, the demand for online access anytime and anywhere as well as new services, such as the multimedia services, results in an enormous expansion of wireless data transmission (traffic), which is continuously
driving the need for capacity increase of wireless communication. From Second-Generation (2G) Global System for mobile communications to Third-Generation (3G) or even the oncoming Fourth-Generation (4G) applications the essential improvement is the spectral efficiency of a wireless communication system measured in bit/s/Hz, besides the increased bandwidth for wireless communication. The new communication systems of 3G and 4G are capable of providing multi-megabits per second throughput by advanced signal processing techniques. For this purpose more and more complex digital modulation schemes are applied in new mobile communication systems, e.g. OFDM or CDMA using Quadrature Amplitude Modulation (QAM). The problem coming with these complex modulation schemes is the high Peak-to-Average Power Ratio (PAPR) of the transmit signals. Furthermore, as the PAs are nonlinear in nature, this not only implies spectral regrowth but also in-band distortions. The transmit signal emitted in licensed frequency bands must meet certain signal integrity requirements, such as Error Vector Magnitude (EVM), Adjacent Channel Leakage Ratio (ACLR) and Spectrum Emission Mask (SEM), depending on specifications for different radio access techniques, e.g. UMTS specific requirements as depicted in Release 6, see [1], [2]. As a consequence of the signal with even higher PAPR, the PAs must be operated with even larger back-off into linear region, which results in even lower PA efficiency. As shown in Fig. 1.2 the PA efficiency for 3G and 4G applications without using of any linearization techniques is obviously lower than for 2G applications due to the increased PAPR.

![Figure 1.2: PA energy efficiency roadmap without using of linearization and reduction of signal dynamic.](image-url)
Because the efficiency theme draws a lot of attention in recent years, the aspect of efficiency becomes more and more important with respect to market requirements and environmental protection policies all over the world. In order to achieve high efficiency, the operating points of PAs must be pushed deeper into their saturation (nonlinear) region. It is generally acknowledged that the PA linearization techniques aiming at the compensation of PA nonlinearities are exactly the solutions to suppress the Adjacent Channel Power (ACP) and reduce the EVM of the transmit signal without lowering the PA efficiency. In contrast to the PA back-off technique, PA linearization techniques allows pushing the operating points of PAs even deeper into their saturation region to achieve even higher efficiency and meeting the required linearity at the same time. The digital predistortion (DPD) is one of the linearization techniques, which has been widely applied in modern wireless communication systems because of its good properties in terms of adaptation possibilities, excellent performance and low implementation effort. With latest DPD technique the average efficiency of conventional PAs can be significantly improved depending on the PA technologies and signal statistics.
Chapter 2 Challenges with Digital Predistortion

Although the DPD techniques have been rapidly improved continuously over the last decades, the demand for a more powerful and robust DPD is growing permanently because of the increased challenging and critical requirements imposed by new applications in wireless communication. Therefore, some critical and open points in DPD technique must be addressed to enable its practical application.

2.1 Research Tasks

In the following paragraphs the main research tasks in this work are presented, including contributions to PA modelling, DPD, and signal conditioning in Envelope Tracking system in terms of compensating the temporal mismatch between the RF signal and its envelope signal.

2.1.1 PA Memory Delay Analysis

It is well known that PAs exhibit some dynamic deviations from their AM-AM and AM-PM characteristics, which can deteriorate the static DPD (memoryless DPD) performance. Such effects are known as “memory effects”, which are understood as an effect where the current PA output signal not only depends on the current PA input signal but also on the history of the input signal. As the signal bandwidth gets wider, the memory effects become more and more significant. Therefore, the DPD also needs to have memory structures in order to compensate the memory effects of PAs. One of the critical problems of memory DPD is to define the time delay constants of memory effects generated by PAs. If these delay constants are known, one is able to improve the DPD performance through optimizing the memory depth of DPD according to these delay constants of PA memory effects. In this work a novel method for delay estimation of the dominating memory effect is developed.
2.1.2 Reliability and Stability of Digital Predistortion

The DPD performance, stability and computational cost are the three aspects, which interact with each other. Generally, the ACLR suppression and EVM reduction increase with the number of coefficients used in DPD. However, a DPD with a large number of coefficients implies at the same time high computational cost by calculation of these coefficients. Furthermore, the instability of the entire DPD system would increase with the number of DPD coefficients as well. In worst case the DPD system would tend to diverge with too many degrees of freedom. For this reason, iterative methods are introduced to find the right DPD coefficients, which exhibit more reliability and stability in comparison to a conventional Least Square (LS) method.

2.1.3 Complexity and Computational Cost of Digital Predistortion

The most powerful DPD with full Volterra representation, see [3], [4], [5], [6], is however not practical because of its enormous number of coefficients. Moreover, the many degrees of freedom also result in instability in practice, as mentioned in the last paragraph. The most popular DPD structures therefore are simplified Volterra DPDs, see [7]. In most cases, the DPD characteristics are constructed by mutually linearly independent functions called basis functions in a certain function space. Theoretically, more basis functions result in a more accurate approximation and a better DPD performance. However, a large number of basis functions also introduce the risk of instability. A DPD designer always needs to find a trade-off between performance, stability and computational cost, so that an intelligent selection of basis functions is very important for a stable and efficient DPD. Instead of blindly choosing basis functions, a novel approach is presented to dynamically select basis functions during the operating time for dynamic reduction of the number of basis functions used at each adaptation step of DPD.
2.1.4 Fast Load Changing of PAs

Another critical issue for DPD is the fast load change in some wireless applications. An OFDM signal with different load cases is depicted in Fig. 2.1. It is obvious that the full load case (blue trace) has the largest dynamic range. The other cases exhibit much smaller dynamic range, depending on the load given in percentages.

This effect of different power range according to various load cases raises a new problem for DPD, which is rarely discussed not only in the context of DPD but also in the context of PA modelling. A PA is usually modelled as a black box by using certain mathematical functions. To extract the coefficients of the mathematical functions, a training signal would be fed through the device under test (DUT). As a consequence, such a model would only hold good for this special training signal. For other signals with different bandwidth or power range, the accuracy of the PA modelling and thus the DPD performance would be most likely downgraded. Moreover, in most common standards the specifications only focus on the full load case, e.g. the test signal model for full-load case is specified for inspection of the ACLR and EVM conformance. Since this issue is not addressed in previous work, a new method is developed in this work to deal with fast load changes.

Figure 2.1: Dynamic load change of OFDM signal.
2.1.5 Temporal Mismatch in Envelope-Tracking System

Conventional RF PAs with constant power supply only offers high efficiency near their maximum output power level. In other words, conventional PAs only work effectively in their saturation region. However, due to the large dynamic range of the transmit signal as shown in Fig. 2.1, the PA efficiency drops significantly, especially for the input signal with very small amplitude. The Envelope Tracking technique is introduced to overcome this efficiency problem through dynamic adaptation of the power supply voltage of PAs to the power of transmit signal, or rather, the power supply voltage would rise and fall following the envelope of the transmit signal. The PAs are therefore always operated in high efficient region independent of the transmit signal power. However, in ET System one has to face another problem, the temporal mismatch between the RF signal and its power supply signal, also know as envelope tracking signal (ET signal). In an ET system if the RF input signal and its dedicated defined power supply voltage do not arrive time aligned at PA, the PA can not provide the designed gain for the current RF signal. As a consequence, the temporal mismatch causes unexpected distortions in an ET system, which would disrupt other compensation processes in PA chain, e.g. DPD and IQ imbalance compensation. This special problem on ET System is analysed in time as well as frequency domain and a compensation approach is also presented in this work.

2.2 Major Contributions

This section summarizes the main contributions of this work in the field of DPD as well as the compensation of temporal mismatch in an ET system.

1. A practical wideband offline DPD test platform was implemented based on a Xilinx FPGA, which enables the hardware experiments of various DPD algorithms.

2. Different DPD structures and algorithms including compensation of PA memory effects were investigated and verified in the abovementioned test platform.

3. A novel approach for delay estimation of the dominating PA memory effect by using Two–Tone test was developed. This approach has much less bandwidth
requirements on the measurement techniques in comparison with other conventional methods.

4. Stationary iterative methods are introduced to find the DPD coefficients, which avoid the inversion of an ill-conditioned matrix and result in a much more stable linearization performance. With proper constraint of the number of iterations one can achieve even less computational cost by using iterative methods in comparison with conventional methods.

5. A novel approach of adaptive selection of basis functions was developed, in which the DPD itself chooses the memory elements and their basis functions to be used in the computation of DPD updates. In this way the computational cost would be significantly reduced without deterioration of DPD performance.

6. A novel approach of attaching old samples was developed to handle fast load changes with offline DPD.

7. A new PA behavioural model for an ET system was proposed to describe conventional PA nonlinearities and the temporal mismatch between RF signal and its ET signal.

8. Temporal mismatch compensation methods for ET system were developed by analysing the feedback signal in time and frequency domain.

2.3 Patents and Publications

Patents


Publications


### 2.4 Literature Review

An extensive study of publications on DPD has been carried out. A good deal of research work on DPD and closely related topics e.g. PA modelling has been done in the last twenty years. Therefore, one can find a large number of publications on different research aspects in terms of DPD structures and algorithms. The detailed list of reviewed publications can be found within List of References.

At first only memoryless nonlinearity of PAs was investigated and characterized by the straightforward and famous polynomial model. Consequently, memoryless linearization methods were developed to fight against the nonlinearity of PAs. Some publications proposed even static DPD methods without adaptation during operating time to simplify the entire DPD system.

With increased bandwidth of transmit signal the performance of memoryless DPD is deteriorated and the conformance with standards as well as the high requirements on PA efficiency cannot be fulfilled simultaneously. The explanation for the signal bandwidth dependent PA nonlinearity is the well-known memory effects. The core idea for compensation of PAs with strong memory effects is that the DPD should have memory effects as well. In the last decade various structures of DPD with memory compensation were investigated. Furthermore, the issues of system stability, implementation effort and computational costs are also addressed in a good number of publications. Table 1 presents an overview of the most relevant publications on DPD with regard to their strengths and weaknesses.
<table>
<thead>
<tr>
<th>Year</th>
<th>Ref. Num</th>
<th>Topic</th>
<th>Strength</th>
<th>Weakness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>[8]-C. Eun, et al.</td>
<td>A predistorter design for a memory-less nonlinearity preceded by a dynamic linear system</td>
<td>+Memory compensation Volterra and Neural Network</td>
<td>-Only simulation results</td>
</tr>
<tr>
<td>2001</td>
<td>[10]-J. Kim, et al.</td>
<td>Digital predistortion of wideband signals based on power amplifier model with memory</td>
<td>+Memory compensation +Individual delay length for memory taps</td>
<td>- No cross terms compensation</td>
</tr>
<tr>
<td>2002</td>
<td>[5]-A. Zhu, et al.</td>
<td>An adaptive Volterra predistorter for the linearization of RF high power amplifiers</td>
<td>+ Memory Compensation with Volterra Structure</td>
<td>-only simulation results -large number of DPD coefficients and high computational costs</td>
</tr>
<tr>
<td>2003</td>
<td>[12]-D. R. Morgan, et al.</td>
<td>Reducing measurement noise effects in digital predistortion of RF power amplifiers</td>
<td>+Dealing with measurement noise</td>
<td>-Solution too complex, not applicable in real system</td>
</tr>
<tr>
<td>2003</td>
<td>[13]-R. Raich, et al.</td>
<td>Digital baseband predistortion of nonlinear power amplifiers using orthogonal polynomials</td>
<td>+Stability improvement through orthogonal basis function</td>
<td>-No individual delay for different memory taps -No cross terms compensation</td>
</tr>
<tr>
<td>2003</td>
<td>[16]-S. Boumaiza, et al.</td>
<td>Thermal memory effects modelling and compensation in RF power amplifiers and predistortion linearizers</td>
<td>+Addressing long term memory effects</td>
<td>-Not addressing short term memory effects</td>
</tr>
<tr>
<td>Year</td>
<td>Authors</td>
<td>Description</td>
<td>Pros</td>
<td>Cons</td>
</tr>
<tr>
<td>-------</td>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
- Turning DPD coefficients according to the power of frequency segments |                                                                      |
| 2007  | [21]-S. Hong, et al.           | Weighted polynomial digital predistortion for Low memory effect Doherty power amplifier | +Using Weighted Polynomial                                   | -Memoryless DPD  
- Using ramp training signal                                         |
| 2007  | [22]-P. L. G. Pinal, et al.    | Multi Look-Up Table digital predistortion for RF power amplifier linearization | +Memory effects depending on PA output | -Risk of instability  
- Additional computational cost for stability test                      |
| 2008  | [4]-A. Zhu, et al.             | Open-Loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra Series | +Reducing the DPD complexity                                       | - No clear metric defined for reduction of DPD complexity             |
| 2010  | [24]-N. Wolf, et al.           | Simple predistortion system for compensation of temperature dependent nonlinearity of power amplifiers | +Introducing temperature dependent DPD coefficients     | -Cannot adapt in operating time  
- Not real time capable                                                  |
| 2010  | [23]-H. Jiang, et al.          | Digital predistortion using stochastic conjugate gradient                  | +Iterative method to solve the coefficients  
+Fast convergence                                                        | -High computational cost  
- No cross terms compensation                                             |

Table 1: Analysis of DPD literatures
Chapter 3 Tools for Assessing Performance

3.1 Simulation Tools

In this work the PA baseband behavioural models are developed by using the simulation tool Matlab from company MathWorks, which is widely employed to simulate and optimize the designs at system level. All the contributions to DPD and ET temporal mismatch compensation described in this work are simulated in Matlab and verified in hardware. Since this work focuses on the compensation of PA nonlinearity, the other analogue imperfections, e.g. IQ imbalance of analogue modulator, are neglected. In hardware experiments the well-designed analogue feedforward and feedback paths exhibit good transfer functions with respect to flatness of frequency response and suppression of undesired spurious signal. The simplification of disregarding analogue imperfections was confirmed to be reasonable, because the observed phenomenon in simulation were obtained in hardware test as well.

The additional benefit of the baseband model simulation is the high simulation speed. The typical PA input signal bandwidth varies from 3.84MHz to 20MHz for single-carrier and multi-carrier UMTS signal, respectively. The baseband simulation is therefore based on sampling frequency of 46.08MHz, 92.16MHz or 184.32MHz in this work. In contrast, the RF band simulation needs much higher sampling frequency, which results in much more data than the baseband simulation in the same time segment and expansion of the simulation time. Furthermore, since the DPD performs in baseband and its coefficients are estimated by the samples in baseband, the PA baseband models are sufficiently accurate to represent PA behaviour.

3.2 Hardware Test Platform

The hardware test platform developed in this work is based on a multi-standard and multi-band capable transmitter, which is being developed in Bell-Labs in Germany. The Fig. 3.1 provides a brief overview of the entire transmitter module. The broadband up-conversion
module and down-conversion module have an operating frequency range from 400MHz to 4GHz. The digital-to-analogue and analogue-to-digital converters are these devices, that convert the digital signal to the analogue signal and vice versa.

Moreover, an analogue I/Q modulator in direct up-conversion transmitter architecture (Zero-IF) exhibits unavoidable mismatches in I- and Q-path, which results in the mirror-frequency interference and carrier leakage. Because these problems can seriously deteriorate the DPD performance, the block of IQ and DC suppression is placed in digital module to compensate imperfection of the analogue I/Q modulator. There are some approaches available to compensate the IQ-impairments, see [27], [28], [29] as well as references therein. But the IQ-Impairments compensation itself is out of concern and will thus not be discussed in the work.

The entire digital signal processing depicted as digital-module in Fig. 3.1 is implemented in a FPGA (Field Programmable Gate Array) evaluation board from Xilinx located on the left side of Fig. 3.2. The board standing in the middle is used to adapt the interfaces of the FPGA evaluation board to the analogue feedforward and analogue feedback boards, which are illustrated on the right side of Fig. 3.2. The small board in the bottom right corner generates the system clock and reference clock for the entire test-setup. A 256MB memory of DDR2 SDRAM plugged in the FPGA evaluation board provide enough space to store an entire frame of test signals with the sampling frequency of 46.08MHz. A UMTS, 3G LTE or WiMAX baseband signal can be pre-downloaded to the DDR2 memory via a PowerPC system. In lab experiments the test signal stored in DDR2 memory would be fed to FPGA continuously and in this way the transmitter can get a constant signal stream.
In order to enable the interactions between user and the hardware as well as debug possibilities a GUI running on a control PC is used to configure the digital and analogue modules, as shown in Fig. 3.3. Common Ethernet connection is applied to carry out the communication between the control PC and the embedded PowerPC system in the FPGA evaluation board.

3.2.1 PowerPC System

The PowerPC system is a low cost high performance embedded system-on-a-chip (SoC). It analyzes the commands received from the Control PC and then sets the control bits in the FPGA or configures the corresponding analogue components via Serial Peripheral Interface
The PowerPC 405 processor in Virtex 4 FPGA has a five-stage pipeline with single cycle execution of most instructions, integrated data- and instruction-caches, core connect bus and other features, see [30]. The benefit of using a PowerPC system in an FPGA is that one can tailor the PowerPC system according to different applications. Moreover, the direct control of the FPGA via a PowerPC system becomes much easier in comparison to a conventional solution with separate FPGA and a separate processor system.

3.2.2 Offline Digital Predistortion of Multi-LUT Structure

In order to test new DPD algorithms with relative low implementation efforts, a flexible offline DPD test platform was developed based on the existing hardware and software. A Multi-LUT (Look-up Table) structure, which can be written and read from a control unit via a PowerPC system, is implemented in the signal path. Since a great number of different DPD algorithms can be covered through various wirings and customisation of these LUTs, one is able to experiment with different DPD approaches of similar structures by writing different contents into these LUTs without extensive modification of the hardware implementation in FPGA. One typical and most well-known DPD structure of parallel LUTs is presented in Fig. 3.4, which is continuously used in the lab experiments in the following chapters.

As mentioned in the previous paragraphs, PowerPC system in the test set-up works as a bridge between the Control PC and the transmitter. The DPD coefficients are calculated in Matlab in the Control PC, then these coefficients would be converted to the contents of the
LUTs and subsequently download to the FPGA. Fig. 3.5 presents a typical process flow chart of the offline DPD. A digital-IO board is used to capture the feedback data down-converted in baseband. Since the original data and the DPD coefficients are known, one is able to calculate the predistorted signal without using the digital-IO board. On the base of predistorted or original data and feedback data the new DPD coefficients are computed in Matlab and subsequently downloaded in the FPGA through the Ethernet connection.

Figure 3.5: Flow chart of offline DPD.
Chapter 4 PA Models and Predistortion Algorithms

The theoretical background of DPD linearization and the novel contributions developed during this doctoral work are presented in this chapter. The first section talks about the PA behavioural modelling with respect to memoryless and memory contained nonlinearity. Subsequently, in Section 4.2 the linearization techniques with particular emphasis on DPD are introduced briefly. Section 4.3 analyzes the Two-Tone experiments and derives the delay constant of the dominating memory effect of PAs. In order to improve the DPD stability and reduce the computational cost simultaneously, the stationary iterative methods and the approach of using dynamical selection of DPD basis functions are presented in Section 4.4, Section 4.5 and Section 4.6. Subsequently, the proposed method of dealing with fast load changes is developed in Section 4.7. Since the ET system stands in the focus of PA design in the last time, the temporal mismatch effects of ET system is addressed in the last section.

4.1 PA Behavioural Models

A linear system has the property, that its output can be expressed as a linear combination (superposition) of responses to individual inputs, e.g. if for inputs $x_1(t)$ and $x_2(t)$, as given by:

$$\begin{align*}
x_1(t) &\rightarrow y_1(t), \\
x_2(t) &\rightarrow y_2(t),
\end{align*}$$

(4.1)

where the arrow indicates the operation of the system, then

$$a \cdot x_1(t) + b \cdot x_2(t) \rightarrow a \cdot y_1(t) + b \cdot y_2(t)$$

(4.2)

for all values of the constants $a$ and $b$.

Any system that does not satisfy this condition is nonlinear. For a nonlinear system the input-output relationship can be approximated with a polynomial, as given by:

$$y(t) = \alpha_0 + \alpha_1 \cdot x(t) + \alpha_2 \cdot x(t)^2 + \alpha_3 \cdot x(t)^3 + \cdots,$$

(4.3)
where $\alpha_j$ are the coefficients, see [31]. A system is time-invariant (TI) if a temporal shift of its input signal results in the same temporal shift at its output. In this work the PAs are assumed to be nonlinear time invariant systems. Generally, the calculation of DPD coefficients is based on the transmitted feedforward and feedback data, which means that the DPD permanently approximates the inverse model of previous behaviour of the PA. Therefore, if PAs would be time variant systems, the DPD approach for compensation of the PA’s previous behaviour would not exhibit good performance in this context. Since the PA is normally regarded as a black box in the PA behavioural modelling, one is free to choose any mathematical model to describe the PA behaviour. In the following paragraphs various PA models with and without memory effects are presented.

### 4.1.1 Memoryless PA Models

The baseband discrete signals, as shown in Fig. 4.1, can be represented as a series of complex number in both a Cartesian coordinate system or a polar coordinate system,

$$x[n] = i[n] + j \cdot q[n]$$

or

$$x[n] = A[n] \cdot e^{j \varphi[n]},$$

where $A[n]$ and $\varphi[n]$ depict the amplitude and phase of the signal $x[n]$, respectively. $x[n]$ and $y[n]$ reflect the complex-valued signal at digital input before up-converter and at digital output after down-converter. Since the Cartesian and polar forms are equivalent, both forms are often used in the derivation of a PA behavioural model.

![Figure 4.1: Simplified transmitter structure.](image-url)
The most well-known and widely used method for PA characterization is the so-called Single-Tone test to measure the amplitude (AM-AM) and phase distortion (AM-PM) with respect to different input power levels. The relationship between the PA input and output signal can be described by:

\[
y[n] = g_r(|x[n]|) \cdot e^{j(x[n]+x[n]|A[n]|)} = g_r(A[n]) \cdot e^{j(x[n]+x[n]|A[n]|)},
\]

where \(g_r(x)\) and \(g_\phi(x)\) are these functions representing the AM-AM and AM-PM characteristics of PAs. The AM-AM characteristic \(g_r(x)\) describes the gain compression or expansion of a PA versus different input power levels at an affixed input frequency. Another important phenomenon, the AM-PM characteristic \(g_\phi(x)\), depicts phase shift of the PA output in comparison to original input signal depending on the input power level as well. It is observed in equation (4.6) that both AM-AM and AM-PM characteristics are functions of the amplitude \(A[n]\) of PA input signal. Thus, the PA nonlinearity depends only on the momentary power of the PA input signal.

As an example, the AM-AM and AM-PM characteristics of a class AB PA are illustrated in Fig. 4.2. Because a Single-Tone test has zero bandwidth, the bandwidth-dependent phenomena can not be obtained in this kind of test.

![Typical AM-AM and AM-PM characteristics of a Class-AB PA.](image-url)
Polynomial Model

As mentioned in a previous chapter, only baseband behavioural models are investigated in this work in order to simplify the PA modelling and accelerate the simulation speed. The polynomial model is often applied to describe the input-output relationship of a PA at baseband in terms of AM-AM and AM-PM characteristics, as described by:

\[
y[n] = \sum_{k=1}^{K} \alpha_k \cdot x[n] \cdot |x[n]|^{K-1},
\]  

(4.7)

where \(\alpha_k\) and \(K\) denote the complex-valued coefficients and polynomial order (degree), respectively. This baseband polynomial model is frequently used to represent the memoryless nonlinear part in a PA memory model.

Saleh’s Model

One of the most commonly used PA memoryless models is the Saleh’s model, because only 4 parameters are required in Saleh’s model to describe the PA behaviour, see [32]. The AM-AM and AM-PM characteristics are expressed separately in two equations

\[
F_{AM}(x[n]) = \frac{\alpha_{AM} \cdot |x[n]|}{1 + \beta_{AM} \cdot |x[n]|^2},
\]

(4.8)

and

\[
F_{PM}(x[n]) = \frac{\alpha_{PM} \cdot |x[n]|^2}{1 + \beta_{PM} \cdot |x[n]|^2}.
\]

(4.9)

Thus, the resultant PA modelled by Saleh’s model can be written as:

\[
y[n] = F_{AM}(x[n]) \cdot e^{j(\phi[n] + F_{PM}(x[n]))},
\]

(4.10)

4.1.2 Memory PA Models

If PAs have only memoryless nonlinearity, the measured AM-AM and AM-PM characteristics should be two solid curves, but instead, one obtains two spreads of AM-AM and AM-PM characteristics in practice, as shown in Fig. 4.2. In other words, each PA input
power value is not only mapped to one PA output power value, but mapped to several
different PA output values. Furthermore, the PA models presented in previous paragraphs
are unable to describe the frequency-dependent PA behaviour, which are defined as memory
effects of PAs. Memory effects can be understood as an effect, where the current PA output
signal depends not only on the current PA input signal but also on the history of the input
signal. In addition, one can also observe asymmetrical Intermodulation Distortions (IMDs)
in the PA output spectrum. Because the classical memoryless PA models have only strictly
symmetrical IMDs in their output spectrum, the observed phenomenon of asymmetrical
IMDs points out again such incompleteness and inaccuracy of these memoryless PA models.

Extensive studies have been made to understand and characterize the memory effects of PAs
in the last decade, see [33], [34]. However, since the causes of memory effects are versatile,
the accurate characterization of PA memory effects is still a very challenging subject.
Generally, the memory effects can be classified into short term and long term memory
effects according to the involved time constants of the memory effects. The short term
memory effects have time constants about several times of carrier signal period. In contrast,
long term memory effects exhibit much lower frequency, typically from a few kHz to MHz
[35]. In the following paragraphs both short and long term memory effects as well as their
possible causes are discussed.

4.1.2.1 Short Term Memory Effects

Wiener Model

The short term memory effects are caused mainly by the reactive components of the active
devices and matching networks [35]. One phenomenon of memory effects is that the IMD
components generated by PAs do not remain constant when varying the tone spacing with a
Two-Tone test signal. Because of the frequency dependent PA behaviour one can directly
think of uneven frequency response of input matching network of PAs. The input matching
network can be represented by an LTI system. As a consequence, this memory PA model
can be represented by cascade of an LTI system and a memoryless nonlinearity (NL), which
is the so-called Wiener model [36], [37], as shown in Fig. 4.3.
The two subsystems are given by:

\[
u[n] = \sum_{q=0}^{Q-1} \gamma_q \cdot x[n - l_q]
\]

(4.11)

and

\[
y[n] = \sum_{k=1}^{K} \alpha_k \cdot u[n] \cdot |u[n]|^{k-1},
\]

(4.12)

where \(\alpha_k\) denotes complex-valued coefficients of the memoryless PA model. \(\gamma_q\) and \(l_q\) describe the impulse response of the input matching network. Substituting equation (4.11) into equation (4.12) gives:

\[
y[n] = \sum_{k=1}^{K} \alpha_k \cdot \left( \sum_{q=0}^{Q-1} \gamma_q \cdot x[n - l_q] \right) \cdot \left( \sum_{q=0}^{Q-1} \gamma_q \cdot x[n - l_q] \right)^{k-1}.
\]

(4.13)

**Hammerstein Model**

Similarly, the PAs have an output matching network as well, which has also certain bandwidth limitation and ripples in its frequency response. A memoryless PA model followed by an LTI system has the name of Hammerstein model [38], as shown in Fig. 4.4.

The two subsystems can be defined by:

\[
u[n] = \sum_{k=1}^{K} \alpha_k \cdot x[n] \cdot |x[n]|^{k-1}
\]

(4.14)

and
\[ y[n] = \sum_{p=0}^{P-1} \beta_p \cdot u[n - l_p], \quad (4.15) \]

where \( \beta_p \) and \( l_p \) describe the impulse response of the output matching network. The relationship between the PA output and input can be expressed as:

\[ y[n] = \sum_{p=0}^{P-1} \beta_p \sum_{k=1}^{K} \alpha_k \cdot x[n - l_p] \cdot |x[n - l_p]|^{k-1}, \quad (4.16) \]

where \( \alpha_k \) represents the complex-valued coefficients of the memoryless PA model.

**Wiener-Hammerstein Model**

If both input and output matching network are taken into account for PA model at the same time, one obtains the so-called Wiener-Hammerstein model, see [39], [40], as illustrated in Fig. 4.5.

\[ x[n] \rightarrow h_1[n] \rightarrow u[n] \rightarrow NL \rightarrow v[n] \rightarrow h_2[n] \rightarrow y[n] \]

*Figure 4.5: The Wiener-Hammerstein model.*

The subsystems are given in the equations (4.11), (4.12) and (4.15), respectively. The output of entire model can be written as:

\[ y[n] = \sum_{p=0}^{P-1} \beta_p \sum_{k=1}^{K} \alpha_k \left( \sum_{q=0}^{Q-1} \gamma_q \cdot x[n - l_p - l_q] \right)^{k-1}. \quad (4.17) \]

**Memory Polynomial Model**

Another popular and useful PA model is the memory polynomial model, see [14], [41], which is constructed by parallel connected memoryless nonlinear models, each with individually delayed input signal. The outputs of the nonlinear sub-models would be summed up, as shown in Fig. 4.6.
The relationship between the input and output of the memory polynomial model can be expressed as:

\[
y[n] = \sum_{q=0}^{Q-1} \sum_{k=1}^{K} \alpha_{k,q} \cdot x[n-l_q] \cdot x[n-l_q]^{k-1},
\]

where \( l_q \) denotes the individual delay for each memoryless subsystem. In rest of this work such memoryless nonlinear subsystem is also termed as function or tap.

**Auto-Regressive Moving Average Model**

Another powerful PA model known as Nonlinear Auto-Regressive Moving Average model is newly developed in [21]. The advantage of this model is to introduce the nonlinear feedback paths in the PA model. In this way the PA current output signal depends not only on the current and previous input data but also on the previous output data of PAs, as shown in Fig. 4.7. However, the nonlinear feedback paths may cause serious stability problem. For this reason, a stability test based on the small-gain theorem needs to be introduced to prevent the divergence of this model. The mathematical representation with polynomial nonlinearity is defined by:

\[
y[n] = \sum_{q=0}^{Q-1} \sum_{k=1}^{K} \alpha_{k,q} \cdot x[n-l_q] \cdot x[n-l_q]^{k-1} + \sum_{p=1}^{P-1} \sum_{j=1}^{J} \beta_{j,p} \cdot y[n-l_p] \cdot y[n-l_p]^{j-1},
\]

where \( \beta_{j,p} \) represents the coefficients of the memoryless functions of input signal \( y[n] \).
Figure 4.7: Auto-Regressive Moving Average model.

**Volterra Model**

Generally the more coefficients a PA model has, the higher the accuracy of this PA model. The most accurate PA model with memory effects is the Volterra model, see [3], [4], [5], [6], [36]:

\[
y[n] = \sum_{l_i=0}^{L} h_k[l_i] \cdot x[n-l_i] + \sum_{k=2}^{K} \sum_{l_i=0}^{L} \cdots \sum_{l_k=0}^{L} h_k[l_i, \ldots, l_k] \cdot x[n-l_i] \cdots x[n-l_k],
\]

(4.20)

where \( h_k[l_i] \) is the discrete Volterra kernels of \( k \)-th order and \( L \) denotes the finite memory length. Usually the Volterra model can achieve higher accuracy in comparison with other PA memory models. However, the number of coefficients of a Volterra model increases exponentially with the memory length and (or) the order of its kernels. Using too many coefficients results not only in high computational cost but also in a decreased stability of the entire system. Because of this weakness the full vision of the Volterra model with deep memory length and high order kernels is unattractive and seldom used in practice. More often, only selected Volterra kernels are used to build a simplified vision of Volterra model. The previously introduced PA memory models, such as the Wiener model and the Hammerstein model, are simplified special cases of the Volterra model.

The Memory Polynomial model is a special case of Volterra model as well, which is composed of the diagonal Volterra kernels. The other Volterra kernels except the diagonal kernels are also called cross terms [7], because these kernels are functions of at least two
successive input samples. In order to improve the performance of the pure Memory Polynomial model, the memory polynomial model is often applied in conjunction with the cross terms.

**Separable Function Model**

The newly developed Separable Function model [20] is actually an alternative representation of the Volterra model, which exhibits similar computational cost and performance as the Volterra model. The output signal of this model can be written as:

\[
y[n] = \sum_{q=0}^{p-1} \prod_{p=0}^{K} \alpha_{k,q,p} \cdot x[n-l_q] \cdot |x[n-l_p]|^{k-1}.
\] (4.21)

### 4.1.2.2 Long Term Memory Effects

In contrast to short term memory effects, long term memory effects are usually caused by dynamic thermal effects of active devices and their biasing networks [42], [43], [44]. In [45] a PA model with long term memory effects caused by a bias network is shown in Fig. 4.8. The relationship between input and output signal is given by:

\[
y[n] = f_{NL}(x[n]) + \alpha_{long} \cdot x[n] \cdot \sum_{q=0}^{Q-1} x^2[n-l_q] \cdot h[l_q],
\] (4.22)

where \( f_{NL}(x) \) presents a memoryless nonlinearity.

![Figure 4.8: PA model with long memory effects.](image)

In [22] the long term memory effects are described by a feedback close-loop, as shown in Fig. 4.9. The PA gain variation can be derived from the dissipated power of PAs with a thermal
low pass filter. Generally the short term memory effects dominate in the PA memory effects, which make the detection and modelling of long term memory effects even more difficult.

\[ y[n] \rightarrow \star \rightarrow PA \rightarrow y[n] \]

\[ 1 \rightarrow + \rightarrow \text{Gain} \rightarrow \text{Thermal Filter} \rightarrow \text{Dissipated Power} \]

*Figure 4.9: PA model with long memory effects using feedback loop.*

**4.2 Power Amplifier Linearization Techniques**

In order to operate the inherently nonlinear PAs with high efficiency and fulfil the requirements on linearity simultaneously, a number of linearization techniques are developed, such as feedback, feedforward and predistortion. These techniques can be used separately or sometimes in combination to compensate the nonlinearities of PAs. In the following paragraphs some of the widely used linearization techniques are briefly presented with respect to their important properties, such as linearization bandwidth and adaptation flexibility, which describes how easily the linearization techniques can detect and compensate characteristic changing of PAs due to temperature drifting, device aging etc..

**4.2.1 Feedback Technique**

In essence the feedback technique described in [46], as illustrated in Fig. 4.10, trades the PA gain for PA linearity and bandwidth. Furthermore, the feedback technique can also make the PA gain more stable to variations in circuit components, i.e. due to temperature effects. Generally the feedback technique is only applied for narrowband applications and has very low flexibility since the analogue components cannot be tweaked during operating time. In contrast to this, the adaptation flexibility of the Cartesian loop feedback, which directly performs the feedback signal on the digital I/Q signal, is much improved.
4.2.2 Feedforward Technique

The principle of feedforward technique is to use an error amplifier to generate the distortion components with 180 degree phase shift to the distortion components generated by main PA, as shown in Fig. 4.11.

In comparison with feedback technique the feedforward technique doesn’t reduce the PA gain and is unconditionally stable without feedback loop [47]. Furthermore, the linearization bandwidth of feedforward technique is very high, up to 60MHz, see [48]. However, the feedforward technique is sensitive to the matching network over the operating bandwidth and its implementation effort is relatively high.

4.2.3 Predistortion Techniques

The intention of predistortion is to generate a distortion block, whose transfer characteristic \( g(x) \) is chosen such that the conjunction of the predistorter (PD) and the PA is a linear system, as depicted in Fig. 4.12.
The spectrum of PD output signal would be at first expanded due to the nonlinear transfer characteristic $g(x)$,

$$x_{\text{DPD}}[n] = g(x[n]).$$

(4.23)

However, the entire system composed of cascaded PD and PA can be looked as a linear system, as given by:

$$y[n] = f(x_{\text{DPD}}[n]) = f(g(x[n])) = K \cdot x[n].$$

(4.24)

where the factor $K$ and function $f(x)$ denotes the desired PA gain and PA transfer function, respectively. The resultant relationship between the normalized transfer characteristics of PD and PA is given by:

$$g(x[n]) = f^{-1}(x[n]).$$

(4.25)

Analogue predistortion [49], [50] and digital predistortion (DPD) [21], [47] are two different linearization techniques depending on how the PD is performed, by analogue signal processing or digital signal processing. Generally the analogue predistortion has the difficulty to compensate the PA memory effects and can hardly be adapted during the operating time. In essence the DPD is better suited for mobile broadband systems [48], since it offers higher PA correction and greater adaptation flexibility at a lower cost. Because the DPD can be
permanently adapted by comparison of the input and output signals of PAs, the characteristic changing of PAs can be easily detected and compensated in time. According to the latest research results the operation bandwidth of a DPD can be extended over 40 MHz due to the improvement of the hardware components, e.g. FPGA and DAC of higher operating frequency.

4.2.4 Digital Predistortion Technique

This work mainly focuses on the DPD, in which the inverse distortion of PAs is estimated by digital signal processing and then performed on the PA input baseband signal. According to the DPD position in the transmitter, there are two different kinds of DPD systems available, namely conventional baseband DPD and Hybrid Digital/RF Envelope predistortion, see [19], as shown in Fig. 4.13 and Fig. 4.14, respectively. The analogue components except PAs are assumed to be perfect and therefore have no impairments on the transmitter.

![Figure 4.13: Conventional DPD in baseband.](image)

![Figure 4.14: Hybrid Digital/RF envelope predistortion.](image)
Since the Hybrid Digital/RF Envelope predistortion is located completely in RF frequency domain, immediately in front of PAs, the predistortion is directly performed on the RF input signal. In contrast, the conventional DPD works in baseband and access the baseband digital I/Q signal instead of the envelope signal. Moreover, the Hybrid Digital/RF Envelope predistortion does not need to know the original I/Q signal and can be therefore used in repeater base station. However, the Hybrid Digital/RF Envelope predistortion has also its own disadvantages, e.g. more hardware resources, such as additional DACs and ADCs.

The DPD discussed in this work is the baseband DPD, see [47], [51], [52], [53], one of the most popular linearization techniques, because of its good properties concerning overall module efficiency, implementation effort and adaptation possibilities. In a baseband DPD system the transmit signal is processed by a PD in the digital baseband domain before it is converted to analogue signal, up-converted to RF, and fed to a PA.

Generally DPD and PA modelling are two very similar applications. The intention of the PA modelling is to find the exact input-output relationship of PAs. In contrast to this, the DPD aims at finding the inverse transfer characteristics of PAs. Even though the tasks of both applications are different, the working principles of both applications are almost identical to each other. They both can be based on the same mathematical models to describe (inverse) nonlinear characteristics of PAs. This is the reason why the most structures of PA modelling mentioned in the previous sections can be used in DPD as well. In the following paragraphs the common DPD structures are presented and briefly discussed in terms of compensation performance and implementation efforts.

At first some useful definitions of technical terms are given, such as direct and indirect learning architecture as well as online (real-time) and offline DPD, which are frequently used through this work.

**Direct and Indirect Learning Architectures**

The baseband model of a transmitter with DPD for both direct learning and indirect learning architecture is depicted in Fig. 4.15. The baseband input signal of PAs (output of DPD) is denoted by \( x_{DPD}[n] \) and the output signal of PAs is described by \( y[n] \). The indirect learning
and direct learning architectures reflect two cases of DPD system. The direct learning architecture compares the original signal $x[n]$ and the normalized feedback signal $y[n]$ to minimize the difference between these two signals. In contrast, the indirect learning architecture tries to directly derive the inverse input-output relationship of PAs through the predistorted signal $x_{\text{DPD}}[n]$ and feedback signal $y[n]$.

![Diagram](image)

**Figure 4.15**: Direct learning and indirect learning architectures of DPD system.

In general indirect learning and direct learning architectures exhibit similar performance, but the direct learning architecture is much less sensitive to the measurement noise in comparison to the indirect learning architecture, since the original input signal $x[n]$ is free from any measurement noise. From the viewpoint of computing aspect the indirect learning architecture is more convenient for offline DPD, because the inverse characteristics of PAs can be directly derived from the predistorted signal $x_{\text{DPD}}[n]$ and feedback signal $y[n]$ by solving a system of equations.

**Online and Offline DPD**

DPD can be categorized into two groups according to how and where the DPD coefficients are estimated. The signal processing of baseband DPD is mostly implemented in FPGA. In contrast, the adaptation algorithm of DPD coefficients can be executed completely within FPGA (at logic circuit level) or separately in a microcontroller system, such as DSP or processor system in FPGA as well.

Online DPD, sometime also known as real-time DPD, is understood as an approach where all the DPD coefficients can be updated per each pair of feedforward and feedback samples. As a consequence, the required hardware resources for an online DPD adaptation algorithm
is extremely low and the complete DPD system including the adaptation algorithm can be implemented in one FPGA device. More accurate to say, the adaptation algorithms of the online DPD are processed at logic circuit level, as shown in Fig. 4.16 (a). The updating speed of online DPD is therefore extremely high and only limited by the sampling frequency of the FPGA device. However, the implementation and debug efforts of online DPD are very high, because the adaptation algorithm must be written in hardware language such as VHDL and Verilog. Furthermore, the online DPD needs some other signal processes in baseband to support its work, e.g. the synchronization and power normalization of the feedforward and feedback data. These signal processes need to be implemented in logic circuit as well. As a consequence, the implementation and debug efforts become even higher.

The alternative of online DPD is the so-called offline DPD, which has been widely used in PA linearization, because it requires less implementation efforts in comparison with online DPD. With offline DPD the coefficients can be calculated separately in a DSP or processors after aggregation of feedforward and feedback data, as shown in Fig. 4.16 (b). The separate DSP and processor system could be a part of a modern FPGA as well. In this way the complex coefficient estimation process including signal synchronization and other signal processes can be easily written in high level language, such as C/C++, instead of a hardware language. The implementation complexity is therefore extremely reduced with offline DPD. In contrast to online DPD, the drawback of offline DPD is the low adaptation speed, because the estimation of new coefficients is based on thousands of feedforward and feedback samples. It takes a long time for the separate DSP or processor system with serial
computation architecture to process matrix operation with a large number of samples. Therefore, the low adaptation speed prevents the practical use of offline DPD for some special applications, in which the load situation could change within micro seconds. The adaptation of the new DPD coefficients cannot follow the load changes, which results in performance degradation of DPD or even divergence of the entire DPD system. This problem would be addressed in Section 4.7 “Fast Load Changing of Power Amplifiers”.

In practice the offline DPD is much more popular for theoretical investigation of different DPD structures and adaptation algorithms than the online DPD concerning the implementation and debug efforts. Through a certain universal hardware test-platform, various DPD structures and algorithms can be experimentally studied without large hardware modifications. This is the reason, why the most DPD hardware experiments are based on offline DPD in this work.

4.2.4.1 Memoryless Digital Predistortion

Just like PA modelling the first generation of DPD was introduced to compensate the pure static nonlinearity of PAs without consideration of memory effects. Theoretically one is able to derive the inverse characteristic of a PA from its AM-AM and AM-PM characteristics. In practice the inverse characteristic can be estimated by using a special training signal, e.g. ramp signal. Under the assumption that the characteristic of the PA remain almost unchanged during the operating time, the DPD coefficients do not need to be adapted. The simplest open loop memoryless DPD is depicted in Fig. 4.17.

![Figure 4.17: Open loop DPD.](image)

The benefit of open loop DPD is that no hardware and software resources are needed for the adaptation algorithm. Unfortunately, the PA characteristic does change during the operating time, depending on the signal statistics, device temperature and so on. The
assumption of no drift of PA characteristic during the operating time is therefore almost inappropriate in practice. In order to adjust the DPD coefficients during the operating time, close loop DPD was developed. As illustrated in Fig. 4.15, the adaptation algorithm compares the feedback data \( y[n] \) with the original \( x[n] \) or the predistorted data \( x_{DPD}[n] \) in the feedforward path. According to the selected pair of data two learning structures for indirect learning and direct learning structures are available for close loop DPD system.

The most popular learning architecture of offline DPD is the indirect learning structure because of its good implementability and straightforward principle. Similar to PA modelling, one just exchanges input signal \( x_{DPD}[n] \) and feedback signal \( y[n] \) of PAs, as shown in Fig. 4.18, the various models and their estimation methods used in PA modelling can therefore be directly applied in offline DPD.

The inverse characteristic of PA can be approximated e.g. through a linear combination of some polynomial basis functions of \( x \cdot |x|^0, x \cdot |x|^1, \ldots x \cdot |x|^{K-1} \). The simplest memoryless DPD based on polynomial approach is given by:

\[
x_{DPD}[n] = \sum_{k=1}^{K} \alpha_k \cdot x[n] \cdot |x[n]|^{k-1}.
\] (4.26)

The DPD coefficients \( \alpha_k \) can be computed by solving the following equations:
\[ x_{DPD}[n] = \sum_{k=1}^{K} \alpha_k \cdot |y[n]|^{k-1}. \quad (4.27) \]

It is worthwhile to point out that the equation (4.27) is very similar to equation (4.26) except replacing the input signal \( x[n] \) by the output signal \( y[n] \). During the operating time the new calculated coefficients \( \alpha_k \) would be fed to the PD continuously. In this way the close loop DPD system can follow the drift of PA characteristic and adjust the inverse curve to achieve good compensation performance.

For instance, a polynomial of 3rd order is used in a DPD system, three coefficients of \( a_1 \), \( a_2 \) and \( a_3 \) need to be found to perform the inverse gain on the input signal. Theoretically one needs only 3 pairs of feedforward and feedback data to calculate the 3 unknowns, as described by:

\[
\begin{align*}
    x_{DPD}[1] &= a_1 \cdot y[1] + \alpha_2 \cdot |y[1]|^{2} + \alpha_3 \cdot |y(1)|^{3}, \\
    x_{DPD}[2] &= a_1 \cdot y[2] + \alpha_2 \cdot |y[2]|^{2} + \alpha_3 \cdot |y(2)|^{3}, \\
    x_{DPD}[3] &= a_1 \cdot y[3] + \alpha_2 \cdot |y[3]|^{2} + \alpha_3 \cdot |y(3)|^{3}. 
\end{align*}
\]  

(4.28)  
(4.29)  
(4.30)

The coefficients can e.g. be solved by using the Gaussian elimination method. In practice thousand pairs of data are aggregated to build an over-determined system of equations because of the following reasons. First of all the three pairs of data represent only three points in the entire DPD curve and only three points can not fully determine a nonlinear transfer characteristic. It is therefore very likely, that the three coefficients are only optimized for these three points in the DPD transfer characteristic and the rest parts of DPD transfer characteristic between or outside the three points may not be properly approximated.

Secondly, the noise effects as well as the impact of single accidental error in a system can be minimized by using a large number of samples. The last reason is the stability issue. It is well known, that the reliability of the solution depends on the condition number of the system of equations. If the selected three pairs of data are very close to each other, the system of equations would become an ill-conditioned matrix. The particular investigation of this
stability issue is addressed later. Usually, polynomials with much higher order are used in DPD to achieve good compensation performance. The over-determined system of equations with \( k \) coefficients and \( N \) sample pairs is given by:

\[
X_{DPD} = M_y \cdot A,
\]

(4.31)

where

\[
X_{DPD} = [x_{DPD}[0], x_{DPD}[1], \ldots, x_{DPD}[N-1]]^T,
\]

(4.32)

and

\[
M_y = \begin{bmatrix}
y[0] & y[0] \cdot y[0]^\dagger & \cdots & y[0] \cdot y[0]^{k-1} \\
\vdots & \vdots & \ddots & \vdots \\
y[N-1] & y[N-1] \cdot y[N-1]^\dagger & \cdots & y[N-1] \cdot y[N-1]^{k-1}
\end{bmatrix},
\]

(4.33)

and

\[
A = [a_1, a_2, \ldots, a_{k-1}]^T,
\]

(4.34)

where \([*]^\dagger\) denotes the transpose of a matrix. The coefficients estimation can be performed offline (e.g. in DSP) by solving an over-determined system of equations by Least Squares (LS) method [54]:

\[
A = \left(M_y^H \cdot M_y \right)^{-1} \cdot M_y^H \cdot X_{DPD},
\]

(4.35)

where \([*]^H\) and \([*]^{-1}\) return the conjugate transpose value and inverse value, respectively.

4.2.4.2 Memory Digital Predistortion

Any PA exhibits some dynamic deviations from its static characteristic, which significantly deteriorate the static DPD performance. It is the PA memory effects explained in previous section of PA modelling. The principle of the memory effects compensation is that the DPD need to have memory effects as well. Therefore, all the PA modelling structures with memory effects can be introduced in the DPD structure for compensation of PA memory effects. Since the various PA memory models have been already extensively explored in
previous paragraphs, the most DPD structures with memory compensation would not be represented here again. As mentioned in the Review of the Literature, various memory DPD structures have their own advantages and disadvantages in terms of compensation performance, system stability, computational costs and implementation efforts. For demonstration purpose only the most widely used DPD structure of memory polynomial is presented in Fig. 4.19.

![Figure 4.19: Indirect learning DPD of memory polynomial structure.](image)

In the adaptation algorithm the update of DPD coefficients $a_{k,q}$ are calculated by exchanging the input and output signal of PAs:

$$\sum_{q=0}^{Q} \sum_{k=1}^{K} a_{k,q} \cdot y[n-l_q] \cdot |y[n-l_q]|^{k-1}, \quad (4.36)$$

where $k$ and $l_q$ indicate the order of the polynomial and the length of memory effects, respectively. In other words, this DPD system with memory compensation is constructed by conjunction of parallel equally or individually delayed memoryless DPDs, as depicted in Fig. 4.19. This memoryless sub-DPDs will be called function or tap later in this work. The coefficients estimation can be done offline (e.g. in DSP) by solving an over-determined system of equations using the LS method, which is given by equation (4.31):
\[ X_{DPD} = M_y \cdot A , \]

where

\[ A = [a_{1,0}, \ldots, a_k, \ldots, a_{k,Q-1}, \ldots, a_{k,Q-1}]^T , \] (4.37)

\[ M_y = [V_{1,0}, \ldots, V_{k,0}, \ldots, V_{1,Q-1}, \ldots, V_{k,Q-1}] , \] (4.38)

with

\[ V_{k,q} = \left[ y[0-l_q], y[0-l_q], \ldots, y[N-l_q], y[N-l_q] \right]^T . \] (4.39)

Similar to memoryless DPD the coefficients \( A \) can be calculated by equation (4.35).

### 4.2.4.3 Online Digital Predistortion

In comparison with offline DPD the adaptation algorithm of online DPD must be much smaller and easier for hardware implementation, because the hardware resource for parallelization of matrix operations with large number of elements is immense. None of the current FPGA devices can provide such large amount of hardware resource. Furthermore, the implementation effort of large matrix operations in VHDL code is often very challenging. Because of these reasons the direct implementation of the offline adaptation algorithm in logic circuit is not feasible. Even if all the difficulties could have been overcome and the offline adaptation algorithm could have been successfully converted into low level VHDL code, the adaptation rate of this solution is still quite low due to the aggregation and processing of such large amount of samples. This kind of solution is therefore not considered as real online DPD.

In contrast to offline adaptation algorithm, the most adaptation algorithms for online DPD are based on single pair of feedforward and feedback samples. As a consequence, the online DPD has an extremely high adaptation rate, theoretically as fast as the sampling clock in the hardware device. Because the online DPD is not the focus in this work, only a simplest example of Look-up Table (LUT) without memory compensation is introduced here. The
schematic of this online DPD is shown in Fig. 4.20. The entire DPD including adaptation algorithm can be easily implemented in a low level logic circuit, e.g. in a FPGA.

![Figure 4.20: Online DPD with LUT structure.](image)

The predistorted signal $x_{DPD}[n]$ is calculated by:

$$x_{DPD}[n] = LUT\{x[n]\} \cdot x[n], \quad (4.40)$$

where the operator $LUT\{m\}$ returns the value stored in the LUT at the address $m$. The residual of current DPD system at each pair of feedforward and feedback sample can be defined by,

$$e[n] = x[n] - y[n]. \quad (4.41)$$

The updating of LUT is given by:

$$LUT\{x[n]\}_{\text{new}} = LUT\{x[n]\}_{\text{old}} + \mu \cdot e[n] \cdot x[n]^*, \quad (4.42)$$

where $[*]$ presents operator of conjugate complex.

In reality the adaptation rate of an online DPD is somehow lower than the sampling clock. The reason is that one does not trust single pair of feedforward and feedback samples, which might be defective due to system error or contaminated by environment noise. Therefore, the averaging mechanism would be applied to average the noise and minimize the impact of system error with a single pair of samples. $\mu$ is the step size that controls the stability and convergence speed of online DPD, usually predefined with a very small value. More algorithms for online (real time) DPD can be found in [55], [56].
4.2.4.4 Digital Predistortion for Mobile Phone Devices

Mobile phone transmitters suffer from the same problem of low efficiency as the base stations because of the less efficient PAs. Furthermore, the battery operating time becomes a very critical issue with the trend towards smaller and lighter mobile phone devices. To increase standby and talk times of mobile phones DPD is a very promising method to reduce PA energy consumption in a mobile phone device. However, due to cost pressure an extra expensive feedback path is not appreciated in most mobile phones. Moreover, the holding of mobile phone can change the antenna impedance, which causes significant distortion in feedback signal and erroneous adaptation of DPD coefficients. Generally, the signal integrity requirements of mobile phone devices are not as stringent as the requirements of base station in terms of ACLR and EVM. Additionally, the low power PAs used in mobile phones exhibit only weak memory effects. Furthermore, the huge extra processing power for memory compensation can be hardly tolerated because of the cost pressure. The low cost open loop DPD is therefore often used in mobile phone devices to improve the PA efficiency. The rest of the work is more focusing on the DPD for base stations with feedback path.

4.3 Memory Delay Estimation by Using Two-Tone Test

The memory effects of PAs deteriorate not only the DPD performance but also the accuracy of PA modelling. Extensive studies on comprehension and compensation of memory effects have been carried out in the last decade. The most well-known phenomenon of the existence of memory effects is the asymmetrical Intermodulation Distortion (IMD) with the PA output spectrum. A conventional memoryless PA model has strictly symmetrical IMDs, which is contradictory to the phenomenon observed in reality. Therefore, the analysis of IMD asymmetry measured by Two-Tone test becomes the common method to characterize the memory effects of PAs. Two-Tone test can be understood as an experiment where only two CW signals at different carrier frequencies in the frequency region of interest are fed to the PA (DUT). This test would be repeated with the two CW signals at various frequencies with different frequency spacing. Thereby, the frequency dependent PA behaviour can be
reflected by the Two-Tone test. In [42], [57] the authors were able to measure the power of IMD3 as well as the relative phase difference between the two IMD3 by applying a cancellation technique. These papers focusing on measurement techniques do not propose a practical PA model to depict the memory effects of PAs.

One critical issue for the characterization of PA memory effects is to find the delay constants of memory effects. Because the interpretations of the memory effects are different in various PA behavioural models, no unified interpretation of the memory delay constants is available, e.g. in the Wiener and Hammerstein model the memory effects are described by an FIR filter in front of or after the static nonlinearity. The delays of the FIR filters reflect the delay constants of PA memory effects. For the most widely used memory polynomial PA model the delay constants of memory effects are defined as the delays of the input signal for each polynomial function.

Unfortunately, there is no very promising method to estimate the time delay constants of PA memory effects. The classic method for weakly nonlinear system tries to analyze the impulse response of PAs and estimates the time delays of memory effects from slope of the impulse signal at PA output, see [35]. Therefore, this classic method has very high bandwidth requirements on the measurement equipments due to generation and measurement of the steep rising and falling edges. Moreover, PAs are no classic linear systems and hence the nonlinearities of PAs may deteriorate the accuracy of the measurement results of these methods for classic linear systems.

In this section a novel method is proposed to estimate the delay constant of the most dominating memory effect based on the asymmetrical IMDs considering the memory polynomial PA model in RF domain. In this method the asymmetrical IMDs are used to estimate the delay constant of the dominating memory effects, which provides the following benefits. First of all, the complete measurements and analysis of Two-Tone test is performed in frequency domain. From the point of view of the bandwidth requirements, the proposed method has very low requirements on the bandwidth of the measurement equipments. Secondly, the nonlinearities of PAs are used to estimate the time delay instead of being
neglected, so that there is no unreasonable simplification, which may downgrade the accuracy of measurement results.

4.3.1 Theory

The deviation of the delay constant of dominating PA memory effect is based on memory polynomial RF model, which can be described by:

\[
y(t) = |a_{1,0}| \cdot \text{real}(e^{j\phi_{1,0}} \cdot \tilde{x}(t)) + |a_{3,0}| \cdot \text{real}(e^{j\phi_{3,0}} \cdot \tilde{x}(t)) + |a_{5,0}| \cdot \text{real}(e^{j\phi_{5,0}} \cdot \tilde{x}(t))^5,
\]

\[
+ |a_{1,1}| \cdot \text{real}(e^{j\phi_{1,1}} \cdot \tilde{x}(t-\tau)) + |a_{3,1}| \cdot \text{real}(e^{j\phi_{3,1}} \cdot \tilde{x}(t-\tau))^3 + |a_{5,1}| \cdot \text{real}(e^{j\phi_{5,1}} \cdot \tilde{x}(t-\tau))^5
\]

where \(|\cdot|\) denotes the operator for absolute value. \(a_{n,m}\) and \(\phi_{n,m}\) describe the complex coefficients and their angles, respectively. The RF model represents the main nonlinearities in a PA as well as the dominating memory effect including AM-AM and AM-PM distortion. \(\tau\) denotes the delay constant of the dominating PA memory function (tap), as shown in Fig. 4.21. Up to fifth odd polynomial is introduced to model the nonlinearity of each tap. The rest of PA memory effects is neglected in this PA model at first.

![Figure 4.21: PA RF memory polynomial model.](image)

The two-tone signal at two frequencies of \(\omega_1\) and \(\omega_2\) can be defined by:

\[
x(t) = \text{real}(A \cdot e^{j\omega_1 t} + A \cdot e^{j\omega_2 t})
\]

\[
\tilde{x}(t) = A \cdot e^{j\omega_1 t} + A \cdot e^{j\omega_2 t}.
\]
where \( x(t) \) represents the real part of \( \tilde{x}(t) \). The complex signal \( \tilde{x}(t) \) is the auxiliary signal for simplification of mathematical derivation. Similarly, the delayed input signal of the memory function is given by:

\[
x(t - \tau) = \text{real} \left( A \cdot e^{j(\omega_1 t - \alpha_1 \tau)} + A \cdot e^{j(\omega_2 t - \alpha_2 \tau)} \right)
\]

(4.46)

\[
\tilde{x}(t - \tau) = A \cdot e^{j(\omega_1 t - \alpha_1 \tau)} + A \cdot e^{j(\omega_2 t - \alpha_2 \tau)}.
\]

First of all, the attention is focused on the 3\(^{\text{rd}}\) order distortion (IMD3) of this PA model, which can be expressed as:

\[
D_{3^\text{rd}} = |a_{3,0}| \cdot \text{real} \left( e^{\frac{\Phi_{3,0}}{3}} \cdot \tilde{x}(t) \right)^3 + |a_{3,1}| \cdot \text{real} \left( e^{\frac{\Phi_{3,1}}{3}} \cdot \tilde{x}(t - \tau) \right)^3.
\]

(4.47)

Because the distortion of 5\(^{\text{th}}\) order (IMD5) is much smaller in comparison to the distortion of 3\(^{\text{rd}}\) order, if PAs do not work in deep nonlinear region, one can assume that the IMD5 and/or the even high order nonlinearity do not have any influence on the 3\(^{\text{rd}}\) order distortion. What counts in a Two-Tone test is only the power of the test tones and their IMDs, the equation (4.47) can be rotated by \(-\Phi_{3,0}\) and simplified to:

\[
D_{3^\text{rd}}^{\text{rotation}} = |a_{3,0}| \cdot \text{real} \left( \tilde{x}(t) \right)^3 + |a_{3,1}| \cdot \text{real} \left( e^{\frac{\Phi_{3,1} - \Phi_{3,0}}{3}} \cdot \tilde{x}(t - \tau) \right)^3.
\]

(4.48)

As is shown in Appendix A, the power of the lower and upper IMD3 can be written as:

\[
\text{IMD3}_\text{L} = \frac{9}{16} |a_{3,0}| \cdot A^6 + \frac{9}{8} |a_{3,1}| \cdot A^6 + \frac{9}{8} |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cos \left( \omega \tau - 3 \Delta \tau + \frac{\Phi_{3,1} - \Phi_{3,0}}{3} \right)
\]

(4.49)

and

\[
\text{IMD3}_\text{U} = \frac{9}{16} |a_{3,0}| \cdot A^6 + \frac{9}{8} |a_{3,1}| \cdot A^6 + \frac{9}{8} |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cos \left( \omega \tau + 3 \Delta \tau + \frac{\Phi_{3,1} - \Phi_{3,0}}{3} \right)
\]

(4.50)

with 

\[
\omega = \frac{\omega_1 + \omega_2}{2}; \Delta = \left| \frac{\omega_1 - \omega_2}{2} \right|
\]
where $\omega$ and $\Delta$ denote the central frequency and half frequency spacing between the two-tone signal, respectively. If $\tau$ equals zero, which indicates a PA model without memory effects, the low and up IMD3 distortions are obviously identical to each other. But because of the memory effect ($\tau \neq 0$) the IMD$_3$-$L$ and IMD$_3$-$U$ become asymmetrical.

In equations (4.49) and (4.50) except the amplitude of test tones $A$, central frequency $\omega$ and half frequency spacing $\Delta$, the other variables are unknown and thus it is impossible to solve the constant $\tau$ with these two equations. However, one is able to cancel the most variables in these equations through the following trick. The difference of the IMD3s is calculated by:

$$IMD \_ \text{diff}(\omega, \Delta) = IMD \_L - IMD \_U$$

$$= \frac{9}{8} \cdot |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cdot \left( \cos \left( \omega \tau - 3 \Delta \tau + \frac{\phi_{3,1} - \phi_{3,0}}{3} \right) - \cos \left( \omega \tau + 3 \Delta \tau + \frac{\phi_{3,1} - \phi_{3,0}}{3} \right) \right)$$

$$= \frac{9}{4} \cdot |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cdot \sin \left( \omega \tau + \frac{\phi_{3,1} - \phi_{3,0}}{3} \right) \cdot \sin(3 \Delta \tau).$$

(4.51)

where $\sin(3 \Delta \tau)$ is the term, which only depends on the frequency spacing of the two-tone signal and the memory delay $\tau$. Therefore, when and only when the central frequency of the two-tone signal remains constant and the frequency spacing of two-tone signal is swept, the last second term $\sin \left( \omega \tau + \frac{\phi_{3,1} - \phi_{3,0}}{3} \right)$ would remains unchanged at all time. Theoretically, one needs only two sets of Two-Tone test result to cancel the most unknowns out and solve the time delay $\tau$, as given by:

$$\frac{IMD \_ \text{diff}(\Delta_1)}{IMD \_ \text{diff}(\Delta_2)} = \frac{\sin(3 \Delta_1 \tau)}{\sin(3 \Delta_2 \tau)} = G.$$  

(4.52)

As can be seen, the equation (4.52) above only depends on the memory delay $\tau$. But this equation is still difficult to solve, because the unknown $\tau$ is given in two trigonometric
functions. In order to further simplify the equation (4.52), in a second step one can simply double the frequency spacing with respect to the frequency spacing used in first step. The formula of time delay $\tau$ can be explicitly expressed as:

$$\frac{IMD_{\text{dif}}(\Delta_1)}{IMD_{\text{dif}}(\Delta_2)} = \frac{\sin(3\Delta_1 \tau)}{\sin(3\Delta_2 \tau)} = \frac{\sin(6\Delta_2 \tau)}{2\sin(3\Delta_1 \tau) \cdot \cos(3\Delta_1 \tau)} = G$$

$$\Rightarrow \tau = \frac{2k\pi \pm \arccos\left(\frac{1}{2G}\right)}{3\Delta_1}; k = \{0, \pm1, \pm2, \cdots\}, \quad (4.53)$$

with $\Delta_2 = 2\Delta_1$.

At first glance, the solution of time constant $\tau$ is not unique. Because PAs are assumed to be a causal system, the time constant $\tau$ must be positive. Furthermore, in hardware test only the time constant $\tau$ with $k = 0$ seems to be realistic and reasonable, so that the resultant equation of time constant $\tau$ is described by:

$$\tau = \frac{\arccos\left(\frac{1}{2G}\right)}{3\Delta_1}. \quad (4.54)$$

In this way, one is able to directly calculate the delay $\tau$ for the dominating PA memory effect from the asymmetry of IMD3. Moreover, not only the asymmetrical IMD3 but also the other asymmetrical IMDs are also observed in Two-Tone test, e.g. IMD5, IMD7, as well as the fundamental tones. Similarly, the memory delay information is hidden behind the asymmetry of all these IMDs and the two fundamental tones, since the dominating PA memory function is approximated by a polynomial of the same degree as the PA main nonlinearity. Under the same conditions of constant central frequency of successive tests and doubled frequency spacing of the second step, a simple and aesthetic mathematical formula is found to describe the relationship between the delay $\tau$ of the dominating PA memory effect and the asymmetrical IMDs, which is given by:

$$\tau = \frac{\arccos\left(\frac{1}{2G}\right)}{n \cdot \Delta_1}, \quad (4.55)$$
where $n$ denotes the order of the IMDs, e.g. 1 for the fundamental tones, 3 for the third order IMD, 5 for the fifth order IMD and so on.

4.3.2 Simulation Results

Simulations are performed in Matlab to validate this idea. A memory polynomial PA model in RF domain with three functions (taps) and $5^{th}$ order polynomial for each tap is used as the test PA:

$$y(n) = \sum_{k=1}^{5} \sum_{q=0}^{2} b_{k,q} \cdot \text{real} \left( x(n - \tau_q) \cdot \frac{Z_{b_{k,q}}}{k} \right)^k. \quad (4.56)$$

The coefficients of the RF PA model,

$$b_{10} = 1.0513 + 0.0904 j, \quad b_{30} = -0.0723 - 0.3868 j, \quad b_{50} = -1.2071 - 0.8787 j,$$

$$b_{11} = -0.0680 - 0.0023 j, \quad b_{31} = 0.2979 + 0.3089 j, \quad b_{51} = -0.3064 - 0.4669 j,$$

$$b_{12} = 0.0289 - 0.0054 j, \quad b_{32} = -0.0828 - 0.1243 j, \quad b_{52} = 0.1536 + 0.1885 j,$$

are derived from the baseband PA memory polynomial model with 3 polynomial functions described in [52], [58]. For the first try only the first two functions are introduced in the simulation to illustrate a PA with only one main function and one memory function, namely the coefficients $b_{k,2}$ are set to be zero. Additive white Gaussian noise (AWGN) is added on the PA output signal to simulate the background noise. The time delay $\tau_0$ and $\tau_1$ are set to be 0 and 10 ns, respectively.

The flow chart shown in Fig. 4.22 represents the simulation process in Matlab. To estimate the memory delay one needs to perform the Two-Tone test twice with same central frequency in one loop. At second test the frequency spacing should be doubled as the frequency spacing used in first test. In this way one is able to use the equation (4.55) to calculate the memory delay. In the following loops the frequency spacing would be increased step by step.
The estimated time delay via the half frequency spacing used in the first test is illustrated in Fig. 4.23. The three red, blue and black curves represent the memory time delay estimated by the power differences between the fundamental tones, IMD3 and IMD5, respectively. All the curves are convergent to the value of 10ns, which confirms to the memory delay introduced in the PA simulation model. Furthermore, one can observe that the time delay estimated by IMD3 with relative high frequency spacing stands much closer to the should-be value of 10ns in comparison to the other curves. Since the power differences of the IMDs are very small for the experiments of narrow frequency spacing and appear as denominator in the equation (4.55), the time delays estimated by large frequency spacing seem to be more reliable.
Figure 4.23: Estimated memory delay constant of PA memory polynomial model with 2 taps.

 Unfortunately, the memory effects of PAs sometimes are quite complicated in reality, which can not be sufficiently described by only one dominating PA memory effect. Therefore, in further simulations a PA memory model with two memory taps are introduced to verify the effectiveness of this method. The entire memory polynomial PA model with 3 taps described in equation (4.56) is used in the simulation. The two time delays $\tau_1$ and $\tau_2$ are set to 10 ns and 20 ns, respectively.

In Fig. 4.24 the three red, blue and black curves present the memory time delay estimated by the power differences of fundamental tones, IMD3 and IMD5, respectively. This method cannot estimate more than one memory delay constant, but one can still observe that the estimated time delay is located between the two time delays $\tau_1$ and $\tau_2$ of this PA model. As a consequence, this proposed method can estimate the time delay of the dominating memory effect of memory polynomial PA model. One can directly use the estimated time delay to construct a memory polynomial PA model with only 2 functions. If the model accuracy is not good enough and one wants to improve the accuracy with more than 2 functions, the estimated time delay can still be considered as a basic reference for selection of the memory delays.
4.3.3 Experimental Results

To verify the effectiveness of this method a Two-Tone test was carried out by the hardware platform described in Section 3.2. The DPD compensation was set to be neutral, so that the two-tone signal can be fed to the PA chain without any modification. The driver ZHL42 from Mini-Circuits was operated in the linear region and hence was assumed to have no nonlinearities and memory effects. The PA used in the experiments was a Class-AB power amplifier with an average output power of 50W. The two-tone signal was modulated at a constant local oscillator (LO) frequency of 2.14 GHz.

As observed in simulation, the memory delay constant estimated by the power differences of IMD3 exhibits better reliability in comparison to the other IMDs. Thereby, in hardware experiments the IMD3 values at PA output are applied to estimate the memory delay constant. The power of each IMD3 can be directly read from the spectrum analyzer FSQ8 of Rohde & Schwarz (R&S). The Video Bandwidth (VBW) and Resolution Bandwidth (RBW) of the FSQ8 is set to be 100 Hz and 10 Hz, respectively, to get a stable and accurate value of each IMD3.

Two sets of the test results at PA full output power level and under the same test conditions are illustrated in Tables 2 and 3. The memory delay constants estimated at relative low frequency spacing are somehow larger than these values estimated at relatively high
frequency spacing. But as observed in simulation, since the delay constants estimated at relative high frequency spacing are more reliable, the memory delay of this PA could be considered as around 5 ns.

<table>
<thead>
<tr>
<th>(f1-f2)/2</th>
<th>Loop2</th>
<th>Loop3</th>
<th>Loop4</th>
<th>Loop5</th>
<th>Loop6</th>
</tr>
</thead>
<tbody>
<tr>
<td>4MHz</td>
<td>-20.18</td>
<td>-22.72</td>
<td>5.29ns</td>
<td>5.54ns</td>
<td>7.29ns</td>
</tr>
<tr>
<td>8MHz</td>
<td>-19.64</td>
<td>-24.17</td>
<td>5.54ns</td>
<td>5.52ns</td>
<td>5.54ns</td>
</tr>
<tr>
<td>4.5MHz</td>
<td>-20.15</td>
<td>-22.78</td>
<td>5.52ns</td>
<td>5.24ns</td>
<td>5.05ns</td>
</tr>
<tr>
<td>9MHz</td>
<td>-19.57</td>
<td>-25.54</td>
<td>5.24ns</td>
<td>5.24ns</td>
<td>5.05ns</td>
</tr>
<tr>
<td>5MHz</td>
<td>-20.08</td>
<td>-23.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10MHz</td>
<td>-19.50</td>
<td>-24.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5MHz</td>
<td>-19.99</td>
<td>-23.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11MHz</td>
<td>-19.46</td>
<td>-25.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6MHz</td>
<td>-19.88</td>
<td>-23.39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12MHz</td>
<td>-19.39</td>
<td>-25.92</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: PA memory delay measurement set 1.

This proposed approach presents the relationship between the well-known asymmetry of IMDs in frequency domain and the PA memory delays in time domain. It is proven by hardware experiments that the time delays estimated by the new method are consistent. From the point of view of the bandwidth requirement, this method has low requirements on the bandwidth of the measurement equipments. Furthermore, the nonlinearities of PAs are used to estimate the time delay instead of being neglected, so that there is no unreasonable simplification, which could deteriorate the accuracy of measurement results. On the other hand this method is sensitive to the hardware platform and the settings of measurements equipments, because the difference of IMDs appears in the denominator and is very small in most case. However, a practical method is provided to estimate the delay of the dominating PA memory effect in this section.

<table>
<thead>
<tr>
<th>(f1-f2)/2</th>
<th>Loop2</th>
<th>Loop3</th>
<th>Loop4</th>
<th>Loop5</th>
<th>Loop6</th>
</tr>
</thead>
<tbody>
<tr>
<td>4MHz</td>
<td>-20.22</td>
<td>-22.72</td>
<td>7.20ns</td>
<td>5.71ns</td>
<td>5.58ns</td>
</tr>
<tr>
<td>8MHz</td>
<td>-19.69</td>
<td>-24.17</td>
<td>5.71ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>4.5MHz</td>
<td>-20.16</td>
<td>-22.79</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>9MHz</td>
<td>-19.58</td>
<td>-25.50</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>5MHz</td>
<td>-20.10</td>
<td>-23.01</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>10MHz</td>
<td>-19.54</td>
<td>-24.92</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>5.5MHz</td>
<td>-19.99</td>
<td>-23.18</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>11MHz</td>
<td>-19.47</td>
<td>-25.41</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>6MHz</td>
<td>-19.90</td>
<td>-23.39</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
<tr>
<td>12MHz</td>
<td>-19.42</td>
<td>-25.89</td>
<td>5.58ns</td>
<td>5.58ns</td>
<td>5.55ns</td>
</tr>
</tbody>
</table>

Table 3: PA memory delay measurement set 2.
4.4 Coefficient Calculation by Using Stationary Iterative Methods

In practice the most widely used structure of offline DPD is the indirect learning structure as shown in Fig. 4.18. The DPD coefficients estimation can be done by solving an over-determined system of equations with LS method [54], as shown in equation (4.35). The entire process seems to be simple and straightforward. However, in equation (4.35) one has to invert a matrix \( (M_y^H \cdot M_y) \), which is ill-conditioned in most cases, especially for the DPD with memory compensation. The stability of a DPD system is dependent entirely on the condition number of the matrix \( M_y \) and the DPD system becomes instable with increased complexity of DPD structure, e.g. when high order Volterra kernels and many memory Volterra kernels are in use.

There are some techniques to increase the stability of DPD systems, i.e. to reduce the matrix condition number. For instance, the orthogonal technique, described in [15], [59], [60] is often utilized to construct a matrix based on orthogonal basis functions for reduction of its condition number. However, the orthogonal techniques can only alleviate this problem of stability, because the orthogonal polynomial DPD with memory compensation still has a matrix with a large condition number. Furthermore, orthogonal technique is not always applicable, since it is difficult to construct orthogonal basis functions for some DPD algorithms. In this section, the estimation of DPD stability with respect to its matrix condition number is presented. Subsequently the stationary iterative methods for solving the DPD coefficients are investigated with respect to their performance and convergence speed, see [61], [62], [63].

The main idea in this chapter is to solve the DPD system of equation not by LS method, but by stationary iterative methods, so that the inversion of an ill-conditioned matrix can be avoided during the computing process. By using the stationary iterative methods the DPD coefficients can be calculated by matrix-vector multiplications and vector additions instead of the matrix inversion used in the equation (4.35). Furthermore, in a special case where an identity matrix is used as a preconditioning matrix for stationary iterative method, one can even avoid all divisions in the entire calculation process.
4.4.1 Stability Estimation of Digital Predistortion System

An indicator for the stability of a DPD system is the condition number of matrix $M_y$. The condition number $\text{cond}(M_y)$ is the ratio of the largest singular value of matrix $M_y$ to its smallest singular value, as described by:

$$\text{cond}(M_y) = \frac{\text{largest singular value of } M_y}{\text{smallest singular value of } M_y}. \quad (4.57)$$

One should consider the condition number as the rough rate, at which the solution $A$ of the system of equations may change with respect to a change in vector $X$:

$$\frac{\|\Delta A\|}{\|A\|} \leq \text{cond}(M_y) \frac{\|\Delta X\|}{\|X\|}. \quad (4.58)$$

Thus, with a large condition number even a small error in $X$ may result in a large error in the solution $A$. To reduce the matrix condition number, the orthogonal technique is often utilized to construct a matrix of orthogonal basis functions. However, the orthogonal technique can only alleviate this problem of high condition number for the DPD with memory compensation. In Fig. 4.25 and Fig. 4.26 the condition numbers of the conventional and orthogonal polynomial presented in [15] for memoryless DPD as well as memory DPD are depicted, respectively. It is worthwhile to point out that the condition number of the matrix $M_y$ not only depends on the DPD structure and algorithm but also on the bandwidth of the transmit signal. Generally, under same conditions the less bandwidth a transmitted signal has, the higher is the condition number of its matrix $M_y$. For both figures a 3-carrier UMTS signal with 15 MHz bandwidth at 184.32 MHz sampling frequency is applied to construct a matrix $M_y$. If one applies other signals with less bandwidth, e.g. single-carrier UMTS signal, multi-carrier GSM signal of lower bandwidth, the condition number would be even larger in comparison to the curves displayed in Fig. 4.25 and Fig. 4.26.
Figure 4.25: Condition number of memoryless DPD with 8096 samples.

The x-axis denotes the highest order of the polynomial in Fig. 4.25. The blue and red curves represent the condition numbers of the conventional polynomial and the orthogonal polynomial, respectively. In the case of memoryless DPD very small condition numbers for orthogonal polynomial are obtained and even the condition number of a DPD with 7 order orthogonal polynomial has a value under 40. In contrast to this, the condition number of a 7 order conventional polynomial is more than 90000.

Figure 4.26: Condition number of memoryless DPD with 8096 samples.

A memory polynomial structure is used to demonstrate the condition numbers for the DPD with memory compensation. In Fig. 4.26 the x-axis denotes the number of memory taps and
each memory function is consisting of a 5\textsuperscript{th} order polynomial. The squares and circles denote the condition numbers of the DPD with 1 clock cycle and 2 clock cycles delay between each memory functions, respectively. Generally, the memory polynomial with short delay distance and numerous memory functions exhibits larger condition number than the memory polynomial with large delay distance and few memory functions, as demonstrated in Fig. 4.26. The condition numbers of orthogonal polynomial, represented by the red curves are much smaller in comparison with conventional polynomial. However, the absolute values of the condition number for orthogonal polynomial are still large, having a value of a few thousands, depending on the polynomial order and delay distance between the memory functions.

In practice, the matrix $M_y$ in equation (4.35) has a condition number normally ranging from several hundreds to several thousands or even higher for orthogonal polynomial DPD with memory compensation. The error range of $\Delta A$ is dependent on the product of the condition number of $M_y$ and the relative vector error $\Delta X$. Assuming that only vector $X$ in the matrix equation has a quantization error of $2^{-17}$ ($\Delta X = 2^{17}$), the error range of the DPD coefficients $\Delta A$ may have a value of $2^7$ with a condition number of 1000. In the worst case the large error of $\Delta A$ can result in a divergence of the DPD system. The cases with less error cause DPD performance variation in terms of ACP suppression and EVM reduction as well as spectral fluctuations during updating of new DPD coefficients. It can be deduced from the results above, that the stability of a DPD system is still a critical problem in practice.

The idea here is to replace the convention LS method by stationary iterative methods for the calculation of DPD coefficients. The stationary iterative methods for solving a linear system of equations embody a quite different approach than direct methods such as an LS method. Stationary iterative methods attempt to solve the system of equations by approaching the solution step by step with an initial guess. The first order and second order stationary iterative methods are investigated in this work. More information about stationary iterative methods can be found in [64].
4.4.2 The First-Order Stationary Iterative Methods

The principle of the first-order stationary iterative methods works as follows. To solve a square system of equations with same number of equations and unknowns, as given by:

\[ X = M \cdot A, \]

(4.59)

the first-order stationary iterative methods start with an initial guess \( A^0 \). The residual \( R^k \), in other words estimation error, can be calculated by substituting the current parameter vector \( A^k \) back to equation (4.59), which leads to:

\[ R^k = M \cdot A^k \cdot X. \]

(4.60)

In each iteration the coefficient vector \( A^k \) is corrected by the vector \( D^{k+1} \), which is calculated out of the residual and a preconditioning matrix \( N \), as described by:

\[ N \cdot D^{(k+1)} = -\tau_k R^k \]

(4.61)

and

\[ A^{(k+1)} = A^k + D^{(k+1)} = A^k - \tau_k (N^{-1} \cdot R^k). \]

(4.62)

\( \tau_k \) is a constant parameter for first-order stationary iterative methods, which influences the convergence rate of first-order iterative methods. The computing process from equation (4.60) to (4.62) would be repeated until one or both break conditions of the following two conditions:

\[ \max\|A^{(k+1)} - A^{(k)}\| < \epsilon_a \]

(4.63a)

and

\[ \max\|M \cdot A - X\| < \epsilon_b \]

(4.63b)

are fulfilled. To stop iterative methods two criteria are available, namely the maximal difference of two successive calculated coefficients \( A^k \) and \( A^{k+1} \) is smaller than a predefined
value $\varepsilon_a$ in equation (4.63a) or the maximal residual of the system of equations (4.59) is smaller than a predefined value $\varepsilon_b$, as shown in equation (4.63b).

However, there is still a hindrance to use stationary iterative methods to solve the DPD coefficients, because the system of equations built for DPD system is an over-determined system of equations instead of a square system of equations. At first the over-determined system of equations must be transformed to a square system of equations without losing information of the originally over-determined system of equations. To overcome this problem one can simply multiply the system of equations (4.59) with the Hermitian transpose of the matrix $M_y$, as given by:

$$M_y^H \cdot (X_{DPD} = M_y \cdot A) \Rightarrow M_y^H \cdot X_{DPD} = (M_y^H \cdot M_y) \cdot A.$$  \hspace{1cm} (4.64)

Based on the equation (4.64) one is able to compute the coefficients of DPD by using the first-order stationary iterative methods. It is worthwhile to point out that a critical task in construction of an efficient first order stationary iterative method is the choice of the preconditioning matrix $N$ and convergence parameter $\tau$, because at each iteration step a new linear system of equations (4.61) with preconditioning matrix $N$ and parameter $\tau$ must be solved. The computational cost of this linear system of equations should be relatively small but still effective for increasing the convergence rate.

The disadvantage of first-order iterative methods is that the rate of convergence may be slow and the stationary iterative methods may even diverge, if the preconditioning matrix $N$ and parameter $\tau$ are improperly selected. Some possible choices of preconditioning matrix $N$ and their optimal parameter $\tau$ for maximal convergence speed are presented in the subsequent paragraphs.

**Basic Stationary Iterative Method without Single Division**

As mentioned previously, one needs to solve the system of equations (4.61) at each iteration step. If an identity matrix is applied as the preconditioning matrix $N$, the system of equations in expression (4.61) becomes the simplest system of equations with minimal computational cost. Furthermore, there is no single division in the parameter estimation
process, which can cause numerical instabilities, if the denominator is extremely small. This iterative method starts from the equation (4.64), with an initial guess of parameter $A^0$. The equation (4.61) can be rewritten as:

$$D^{(k+1)} = -\tau \cdot R^k.$$  \hspace{1cm} (4.65)

The DPD coefficients are calculated only by matrix multiplications and additions, as given by:

$$A^{(k+1)} = A^k + D^{(k+1)} = A - \tau \cdot R^k. \hspace{1cm} (4.66)$$

The optimal parameter $\tau$ for maximal convergence rate can be calculated according to:

$$\tau_{opt} = \frac{2}{(\lambda_{min} + \lambda_{max})}, \hspace{1cm} (4.67)$$

where $\lambda_{min}$ and $\lambda_{max}$ are the minimal and maximal eigenvalues of matrix $\left(M^H_y \cdot M_y\right)$, see [64]. For the basic stationary iterative method one needs to calculate the extreme eigenvalues of matrix $\left(M^H_y \cdot M_y\right)$, which causes additional computational cost. In the following paragraphs two other first-order stationary iterative methods are introduced, by which the extreme eigenvalues do not need to be calculated upfront.

**Jacobi Method**

In Jacobi method the matrix $\left(M^H_y \cdot M_y\right)$ in equation (4.64) is decomposed in three matrices $L$, $D$ and $U$, that represent the diagonal, strictly lower triangular and strictly upper triangular parts of $M^H_y \cdot M_y$, as given by:

$$\left(M^H_y \cdot M_y\right) = L + D + U \hspace{1cm} (4.68)$$

with

$$L = \begin{bmatrix} 0 & 0 & 0 & \cdots \\ \tilde{y}_{31} & 0 & 0 & \cdots \\ \tilde{y}_{32} & \tilde{y}_{32} & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad D = \begin{bmatrix} \tilde{y}_{11} & 0 & 0 & \cdots \\ 0 & \tilde{y}_{22} & 0 & \cdots \\ 0 & 0 & \tilde{y}_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \quad U = \begin{bmatrix} 0 & \tilde{y}_{12} & \tilde{y}_{13} & \cdots \\ 0 & 0 & \tilde{y}_{23} & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \hspace{1cm} (4.69)$$
The diagonal part $D$ is used as preconditioning matrix $N$ and the parameter $\tau$ is set to be one. In this way one does not need the extreme eigenvalues of matrix $\left(M_y^H \cdot M_y\right)$ to find the optimal parameter $\tau$, but instead of that the convergence condition of Jacobi method must be verified, i.e. the spectral radius of iteration matrix $D^{-1}(L+U)$ needs to be smaller than one [64]. The single coefficient $a_i^{k+1}$ comprised in the vector $A^k$ is calculated by:

$$a_i^{k+1} = \frac{1}{\tilde{y}_{ii}} \cdot \left(-\sum_{j=1 \atop j \neq i}^n \tilde{y}_{ij} \cdot a_j^k + \tilde{x}_i\right). \quad (4.70)$$

The benefit of the Jacobi method is already mentioned, that the computational cost for calculation of extreme eigenvalues can be saved. But on the other side one needs to verify the convergence condition of Jacobi method, which means additional computational cost as well. If the matrix $\left(M_y^H \cdot M_y\right)$ is strictly diagonal dominant, the Jacobi method would always converge. But since the transmit signals are almost random signals in practice, the matrix $\left(M_y^H \cdot M_y\right)$ is obviously not strictly diagonal dominant and thus one needs to verify the spectral radius of the iteration matrix $D^{-1}(L+U)$ before using Jacobi method.

**Gauss-Seidel Method**

The Gauss-Seidel method is actually derived from Jacobi method. The basic idea of Gauss-Seidel method is to use the most up-to-date coefficients value $a_i^{k+1}$ for calculation of the remainder $a_i^{k+1}$. In this method the matrix $\left(L + D\right)$ is used as preconditioning matrix $N$ and the single coefficient $a_i^{k+1}$ is computed by:

$$a_i^{k+1} = \frac{1}{\tilde{y}_{ii}} \cdot \left(-\sum_{j=1}^{i-1} \tilde{y}_{ij} \cdot a_j^{k+1} - \sum_{j=i+1}^n \tilde{y}_{ij} \cdot a_j^k + \tilde{x}_i\right). \quad (4.71)$$

The convergence condition of Gauss-Seidel method is more relaxed compared to Jacobi method. If the matrix $\left(M_y^H \cdot M_y\right)$ in the equation (4.64) is Hermitian positive-definite, the Gauss-Seidel method would converge [65] and no extra convergence condition needs to be
approved before applying the Gauss-Seidel method. Fortunately the matrix \((M_y^H \cdot M_y)\) is always Hermitian positive-semi-definite, independent of DPD algorithms and transmit signals. The detailed proof is given in Appendix B. Although the convergence condition of Gauss-Seidel method is stricter than the character of the matrix \((M_y^H \cdot M_y)\), the Gauss-Seidel method always converges in simulations and hardware experiments. It is very unlikely that the matrix \((M_y^H \cdot M_y)\) is zero-definite. In other words no request on verifying of the convergence condition is actually needed for using the Gauss-Seidel method.

4.4.3 The Second-Order Stationary Iterative Methods

By second-order iterative methods the new set of coefficients is calculated on the base of last two sets of coefficients, as defined by:

\[
A^{(k+1)} = A^k + D^{(k+1)} = \alpha_k \cdot A^k + \left(1 - \alpha_k\right) \cdot A^{k-1} - \tau_k \cdot (N^{-1} \cdot R^k),
\]

where \(\alpha_k\) and \(\tau_k\) denote the two constant parameters for all \(k\) by stationary iterative methods. Similar to the first order stationary iterative methods, the critical task in the construction of an efficient second-order stationary iterative method is the choice of preconditioning matrix \(N\) as well as convergence parameters \(\alpha\) and \(\tau\). For a certain preconditioning matrix \(N\) the optimal convergence parameters \(\alpha\) and \(\tau\) are defined by:

\[
\alpha_{opt} = 1 + \left(\frac{1 - \sqrt{\lambda_{\min} / \lambda_{\max}}}{1 + \sqrt{\lambda_{\min} / \lambda_{\max}}}\right)^2,
\]

and

\[
\tau_{opt} = 2 / (\lambda_{\min} + \lambda_{\max}),
\]

where \(\lambda_{\min}\) and \(\lambda_{\max}\) denote the minimal and maximal eigenvalues of matrix \((N^{-1} \cdot M)\) [64].

Generally in comparison to first-order stationary iterative methods the second-order stationary iterative methods have a higher convergence rate. In simulation and hardware experiments the convergence rate of second-order stationary iterative methods are almost ten
times higher than convergence rate of first-order stationary iterative methods with same preconditioning matrix $N$ and optimal convergence parameters. But the drawback of second iterative methods is the additional computational cost for calculation of the parameters $\alpha_{opt}$ and $\tau_{opt}$.

Among the stationary iterative methods the Gauss-Seidel method is a good choice in comparison to the other first-order and second-order stationary iterative methods. The key advantage of the Gauss-Seidel method is no need for computation of extreme eigenvalues and verifying of its convergence condition. The convergence rate of the Gauss-Seidel method is also relatively high, which depends again on the condition number of $M_y$.

### 4.4.4 Simulation Results

Generally the number of iterations for iterative methods is strongly dependent on the convergence rate, the break conditions and initial values. To accelerate the computation, the old DPD coefficients can be used as the initial values for adaptation of the new coefficients. Furthermore, one can relax the break conditions to achieve an even less number of iterations or simply limit the maximal number of iterations.

![Gauss-Seidel Method](image)

**Figure 4.27:** DPD performance vs. maximal number of iterations for Gauss Seidel method.
The simulation results for DPD performance vs. maximal number of iterations are presented in Fig. 4.27. A baseband PA model of Volterra structure is used as test object and compensated by an orthogonal memory polynomial DPD. The Gauss-Seidel method is used to solve the DPD coefficients with a set of neutral initial values, i.e. the DPD has a constant gain 0dB for all input power level. It can be observed that the ACP is more suppressed with increased number of iterations. However, the difference of DPD performance between 30 iterations and 300 iterations is almost negligible small.

A similar simulation is carried out for DPD performance vs. maximal residual, as shown in equation (4.63a). The smaller the maximal residual is, the better performs the DPD in terms of the suppression of ACP. It is obvious that the break-conditions of maximal residual and number of iterations correlate with each other, namely a small residual value indicates a large iteration number and vice versa. The trade-off between the break-conditions in consideration of computational cost and DPD performance can be optimized in practice. The over stringent break conditions cause only high computational cost instead of significant improvement of DPD performance.

A similar simulation performance is also observe by using of other first order and second order stationary iterative methods. Furthermore, does one really need a completely new set of DPD coefficients to replace the old one, even when the old coefficients exhibit good ACP reduction and EVM correction? The answer is “not required”, because one does not know, if
the completely new calculated coefficients are 100 percent correct or not. Generally the new DPD coefficients calculated by direct methods e.g. LS method have no consideration of the old DPD state. The difference between the old coefficients and new coefficients may be very large, if the condition number of DPD system is very large or the captured feedforward and feedback samples have certain power normalization mismatch or synchronization misalignment. If the new calculated coefficients are directly applied in a DPD system, the PA output spectrum would have strong fluctuations during updating of the new coefficients, which would probably violate the specification of ACP. In practice the new applied coefficients therefore partially consist of the old coefficients and the new estimated coefficients, as given by:

\[ a_{\text{used,new}} = (1 - \xi) \cdot a_{\text{used,old}} + \xi \cdot a_{\text{new}} \]  \quad (4.75)

where \( \xi \) denotes the weighting factor of the new calculated DPD coefficients. In contrary to direct methods, the stationary iterative methods have a direct and strong connection to the old DPD state. The old DPD coefficients can be used as the initial value for coefficients adaptation in stationary iterative methods. In addition, one is able to stop the iterative methods at any time through the definition of the break conditions or restriction of the maximal iteration number. In this way one can easily replace equation (4.75) in stationary iterative methods with early stopping, which can not only reduces the computational cost but also improves the stability of the entire DPD system.

### 4.4.5 Experimental Results

The DPD system with memory compensation was implemented in an FPGA, as presented in the chapter of the hardware test platform. A Class-AB power amplifier with an average output power of 50W was tested by using different UMTS signals. The orthogonal polynomial is applied to construct a memory polynomial DPD, which’s coefficients are computed in Matlab by various iterative methods. Subsequently, the coefficients are converted into LUTs and downloaded to the signal path implemented in FPGA.
Fig 4.29 presents the PA output spectrum with memory polynomial DPD using Gauss-Seidel method. It is evident that the DPD works very well in conjunction with the stationary iterative method (Gauss-Seidel method). The ACP power is suppressed over 20dB and the EVM value is reduced from 11.5% to 8%. The rest 8% EVM is mostly caused by PAPR reduction process and imperfection of the other analogue components, which can not be compensated by DPD.

It is proven in lab experiments that DPD solving by iterative methods exhibits exactly the same performance as the DPD solving by conventional direct method in stable state. But the stationary iterative methods provide much more stable performance during updating of the new coefficients.

For comparison purpose, the DPD is performed in experiments by using the LS and Gauss-Seidel method. The DPD characteristic curves of main function and first memory function are illustrated in Fig. 4.30 for a single-carrier and in Fig. 4.31 for a 3-carrier UMTS test signal. For LS method all curves of first memory functions are far-scattered, especially in AM-PM distortion, which results in strong spectral fluctuation (up to 20dB in ACP region) of the PA output during updating coefficients. For the Gauss-Seidel method (stationary iterative method) the curves of each function are bundled up. One observes only a slight spectral fluctuation (below 3dB in ACP region) of PA output during the update by new coefficients. Moreover, since the spectral fluctuation strongly depends on the iteration number of Gauss-
Seidel method, one can reduce the maximal number of iterations to guarantee even smoother updating of new coefficients. Other iterative methods, such as basic first order stationary iterative method, Jacobi method and the second-order stationary iterative methods, exhibit similar performance to the Gauss-Seidel method.

Figure 4.30: Orthogonal polynomial DPD characteristics of LS method and Gauss-Seidel method for a single-carrier UMTS signal.

Figure 4.31: Orthogonal polynomial DPD characteristics of LS method and Gauss-Seidel method for a 3-carrier UMTS signal.

The coefficients of a DPD system solved by stationary iterative methods are much more stable in comparison to those solved by the LS method, because the inversion of an ill-conditioned matrix can be avoided in stationary iterative methods. The static performance of DPD would not be deteriorated by using the stationary iterative methods. In addition the old DPD coefficients can be used as initial values for the coefficient adaptation. As a consequence the computational cost would be significantly reduced. In stationary iterative
methods one has complete control on coefficients computation, e.g. stopping iterative methods at any time according to the different system requirements. Furthermore, one can even adaptively change the break condition of the stationary iterative methods with respect to the requirements on EVM or ACP. If the EVM or ACP is small enough, one can relax the break condition to reduce the iterations. On the other side, if the EVM value or ACP value is too large, one can make the break condition more stringent to increase the iterations and thus the DPD would have better performance. In this way one is able to have full control of the DPD in terms of compensation performance, adaptation speed and updating step size of DPD coefficients.

4.5 Digital Predistortion Using Adaptive Selection of Basis Functions

Generally the more complicated the DPD structure is, the better the DPD performs. Unfortunately a high degree of freedom for DPD also results in instability at the same time. The most powerful DPD with full Volterra representation is therefore not practical because of its enormous number of coefficients and hence the most popular DPD structures are simplified Volterra structures. Thus, one needs to choose the Volterra kernels to build an individual DPD structure for different PAs. Typically the DPD characteristic is the linear combination of some linearly independent basis functions like kernels in Volterra structure in a certain DPD function space. Let \( \{bf_0(x), bf_1(x), \ldots, bf_{m-1}(x)\} \) be a set of linearly independent complex valued basis functions of one real variable defined on the interval \([0, 1]\). A memoryless DPD function can be expressed as:

\[
    f_{DPD}(x) = \sum_{k=1}^{K} a_k \cdot bf_k(|x|). \tag{4.76}
\]

For instance, a typical polynomial space of degree \( K \) contains basis functions of \( x^0, x^1, \ldots, x^{K-1} \). Theoretically more basis functions enable more accurate approximation and hence a higher linearization performance. However, more basis functions also introduce more computational cost and worse stability of DPD. A DPD designer always tries to find a balance between performance, stability and computational cost. On this account, an
intelligent selection of DPD structure and DPD basis functions is very important for DPD design.

Without the knowledge of which basis functions could make more contribution to compensation of PA nonlinearities, one just chooses the first several basis functions in a certain DPD space. If memory compensation is required, extra memory functions consisted of the same basis functions are involved in the DPD. Without prior knowledge of PA characteristics, one often tends to use maximum memory functions and highest degree of basis functions, as hardware/software resources can allow, in order to achieve the best possible DPD performance. A large number of memory functions composed of high order basis functions however often leads to an ill-conditioned system of equations, as illustrated in Fig. 4.25 and Fig. 4.26. The DPD would suffer from instability, if a large number of memory functions and basis functions are used. Furthermore, the complexity and computational cost is increased if more memory functions and basis functions than necessary are in use. Otherwise, one can try to identify DPD memory functions and try out all possible combinations of these available basis functions. First, this method must be repeated for different PAs. Second, the aging and temperature drift can change the characteristics of PAs, which may result in requirements on new memory functions and/or basis functions to compensate the PA nonlinearities.

Ideally, the contribution of each basis functions to the DPD performance in terms of ACP suppression and EVM reduction can be identified during the PA operating time. The memory functions and basis functions should be selected dynamically and they should be added to or removed from the DPD space adaptively according to whether they can contribute to improving the DPD performance [66]. In other words, the DPD can decide which memory functions (taps) with which basis functions are considered in construction of the DPD. Therefore, in this section a novel DPD approach of “using adaptively selected basis functions during DPD operating time” is presented.

The most basis function based DPD structures can use this proposed method to dynamically select the basis functions. For demonstration purpose, the well-known memory polynomial DPD structure is given in form of basis function representation, as given by:
where \( y[n] \) and \( x_{DPD}[n] \) denote the output and input signal of PA, respectively. The parameters \( k \) and \( q \) present the polynomial order and the delay length of the memory functions. The basis function \( bf_{k,q}(x) \) of the conventional polynomial and the orthogonal polynomial presented in [15] are defined by:

\[
bf_{k,q}(y[n]) = y[n - d_q] \cdot |y[n - d_q]|^{q-1}
\]

(4.78)

and

\[
bf_{k,q}(y[n]) = y[n - d_q] \cdot \sum_{l=1}^{k} (-1)^{l+1} \cdot \frac{(k + l)!}{(l-1)!(l+1)!(k-l)!} \cdot |y[n - d_q]|^{q-1}.
\]

(4.79)

respectively. The DPD function space is therefore composed of \( K \cdot Q \) basis functions, as given by:

\[
S = [bf_{1,0}, bf_{2,0}, \ldots, bf_{k,0}, bf_{1,1}, \ldots, bf_{k,Q-1}].
\]

(4.80)

In this approach a metric for each basis function is introduced to measure how much each basis function contributes to improvement of DPD linearization performance. The metric will be used to select the relatively “important” basis functions and the selected basis functions will be applied to build the function space for DPD adaptation. Moreover, in order to effectively minimize the number of basis functions only the coefficient updates are computed at each iteration. As a consequence, the coefficients updates are added to the old DPD coefficients of the previous iterations. This approach of using adaptively selected basis functions during DPD operating time can be used in conjunction with both direct and indirect learning DPD architectures. In the following paragraphs the approach is separately presented for both these DPD architectures.
4.5.1 Indirect Learning DPD Using Adaptive Selection of Basis Functions

The equation (4.77) is a typical form of the indirect learning DPD architecture, because the system of equations contains the predistorted signal $x_{\text{DPD}}$ and the feedback signal $y$. The metric of “importance” of each basis function is the correlation between the basis function and the residual of the DPD system of equations, as defined in:

$$\gamma_{k,q} = \sum_{q=0}^{Q-1} \left( b_{k,q} \cdot (y[n])^\ast \cdot (x_{\text{ref}}[n] - \tilde{x}_{\text{ref}}[n]) \right)$$ \hspace{1cm} (4.81)

where

$$\tilde{x}_{\text{DPD}}[n] = \sum_{q=0}^{Q} \sum_{k=1}^{K} a_{k-1,q} \cdot b_{k,q} \cdot (y[n])$$ \hspace{1cm} (4.82)

and

$$x_{\text{DPD}}[n] = \sum_{q=0}^{Q} \sum_{k=1}^{K} a_{k-1,q} \cdot b_{k,q} \cdot (x[n])$$ \hspace{1cm} (4.83)

The feedback signal $y[n]$ is substituted back into the DPD system with current DPD coefficients to calculate a predicted value of the predistorted signal $\tilde{x}_{\text{DPD}}[n]$, as shown in equation (4.82). $x_{\text{DPD}}[n]$ is the captured predistorted signal, which can also be calculated by the equation (4.83). The difference of predicted value $\tilde{x}_{\text{DPD}}[n]$ and the captured $x_{\text{DPD}}[n]$ is defined as the residual of the indirect learning DPD system. The correlations of all basis functions are given in

$$r = [\gamma_{0,1}, \gamma_{0,2} \cdots \gamma_{K,Q}]$$ \hspace{1cm} (4.84)

With the help of two predefined or dynamically optimised threshold values $\delta_1$ and $\delta_2$, one is able to evaluate, how large the basis functions correlate with the residual of DPD system. Therefore, one is able to add any “important” basis functions into the current DPD space $S$ or remove any “unimportant” basis function from the space $S$, as described by:
\[ S \leftarrow S \cup \{ bf_{q,k} \} \text{ if } bf_{q,k} \notin S \text{ and } \gamma_{q,k} > \delta_1 \]  

\[ (4.85a) \]

and

\[ S \leftarrow S \setminus \{ bf_{q,k} \} \text{ if } bf_{q,k} \in S \text{ and } \gamma_{q,k} < \delta_2 . \]  

\[ (4.85b) \]

The DPD space \( S \) consists of these basis functions that are determined to be “important” towards the improvement of DPD performance, because the residual has strong correlation with these basis functions. The basis functions not contained in \( S \) are considered as “unimportant”, because the residual has small projections on these basis functions. In this way only the basis functions in \( S \) will be used in the computation of the updates of DPD coefficients and other basis functions will not participate in the computation.

Another improvement of this new approach is that instead of completely new DPD coefficients only the DPD coefficient updates are calculated. If completely new DPD coefficients are calculated like conventional existing solutions, one can only add new basis functions into the DPD space \( S \). Because the basis functions in \( S \) have already made contributions to linearization, the correlation between these basis functions and the residual may be very small. When the completely new DPD coefficients are calculated, one must keep these “old” basis functions, which already existed in the DPD space \( S \). Therefore, in the new approach only DPD coefficient updates are computed and these basis functions with small projections from the DPD residual can be removed from the DPD space.

The new coefficient vector \( A \) in equation (4.31) can be split into the current coefficient vector \( A^{\text{old}} \) and the updates \( \Delta A \), as given by:

\[ X_{\text{DPD}} = M_y \cdot (A^{\text{old}} + \Delta A) . \]  

\[ (4.86) \]

Subsequently, the term \( M_y \cdot A^{\text{old}} \) is moved to the left side of the equation (4.86) and the resultant new system of equations is expressed as:

\[ X_{\text{DPD}} - M_y \cdot A^{\text{old}} = M_y \cdot \Delta A . \]  

\[ (4.87) \]
It is worthwhile to point out that the matrix $M_y$ in equations (4.86) and (4.87) consists of all basis functions that have belonged to the DPD space $S$ at the earlier iterations. Similarly, the solution $A^{old}$ is a vector in the span of all basis functions that have belonged to the DPD space at the earlier iterations. The matrix $M_{bf}$ is composed of the basis functions currently existing in the DPD space $S$. The $\Delta A$ is a vector in the span of these basis functions used in the matrix $M_{bf}$. The new DPD coefficients can thus be calculated by:

$$A^{new} = A^{old} + \Delta A.$$  

(4.88)

Thereby, one is able to only update these DPD coefficients, in which the DPD residual has a large projection. The rest of the DPD coefficients, that have small projections, do not need to be modified at all.

4.5.2 Direct Learning DPD Using Adaptive Selection of Basis Functions

As mentioned before, this approach can be used for the direct learning DPD architecture as well. In the direct learning DPD the residual of the DPD system can be simply defined as the difference between the original signal $x[n]$ and the feedback signal $y[n]$, so that the correlation between the DPD system residual and the basis functions can be presented by:

$$\gamma_{q,k} = \sum_{p=0}^{N} (b_{f,k} \cdot (x[n] - y[n])) = \gamma_{0,k}.$$  

(4.89)

If the DPD works correctly, the feedback signal should be identical with the original signal. Consequently, one can even replace the feedback signal $y[n]$ by the original signal $x[n]$ as the input signal of the basis functions, which leads to

$$\gamma_{q,k} = \sum_{p=0}^{N} (b_{f,k} \cdot (x[n] - y[n])) = \gamma_{0,k}.$$  

(4.90)

Under the condition of perfect linearization these two equations (4.90) and (4.91) are completely equivalent. But the equation (4.90) is more preferred in practice. First of all, the
original signal $x[n]$ comes directly from digital baseband and is therefore almost free from any kind of interference and measurement noise. In contrary, the feedback signal $y[n]$ is down converted and digitalized from analogue signal and hence it is very likely that the feedback signal is accompanied by some interference and measurement noise. Furthermore, the huge benefit of equation (4.90) in comparison to equation (4.89) is that it has much lower computational cost, because the basis functions $bf_{q,k}(x[n])$ of input signal $x[n]$ have been already calculated in PD. Therefore, these terms can be temporally stored for the later use in equation (4.90). The new system of equations for calculation of DPD updates is given by:

$$X - Y = M_{bf} \cdot \Delta A,$$

with

$$M_{bf} = \begin{bmatrix}
bf_{1,0}(y[n]) & bf_{2,0}(y[n]) & \cdots & bf_{K,Q}(y[n]) \\
bf_{1,0}(y[n-1]) & bf_{2,0}(y[n-1]) & \cdots & bf_{K,Q}(y[n-1]) \\
\vdots & \vdots & \ddots & \vdots \\
bf_{1,0}(y[n-N-1]) & bf_{2,0}(y[n-N-1]) & \cdots & bf_{K,Q}(y[n-N-1])
\end{bmatrix},$$

or

$$M_{bf} = \begin{bmatrix}
bf_{1,0}(x[n]) & bf_{2,0}(x[n]) & \cdots & bf_{K,Q}(x[n]) \\
bf_{1,0}(x[n-1]) & bf_{2,0}(x[n-1]) & \cdots & bf_{K,Q}(x[n-1]) \\
\vdots & \vdots & \ddots & \vdots \\
bf_{1,0}(x[n-N-1]) & bf_{2,0}(x[n-N-1]) & \cdots & bf_{K,Q}(x[n-N-1])
\end{bmatrix}. $$

Similarly, the new DPD coefficients are calculated by the equation (4.88). The matrix $M_{bf}$ composed of selected basis functions again has two expressions of equation (4.92) and (4.93), according to the two correlation metrics of equation (4.89) and (4.90), respectively. It is worthwhile to point out that the terms of basis functions in equation (4.93) have been already calculated in PD, so that a large amount of the computational cost can be saved.

Generally this proposed method of the adaptive selection of basis functions is favoured to be applied in conjunction with the DPD of direct learning architecture, because it has much less
computational cost and much less sensitivity to noise and interference than in case of the indirect learning architecture.

### 4.5.3 Simulation Results

As usual, simulations are performed in a Matlab environment with a polynomial PA model containing memory effects. The direct and indirect learning architectures with orthogonal basis functions are tested. The orthogonal basis \( \{bf_0(x),...,bf_{M-1}(x)\} \) presented in [18] can be constructed in a simpler way by using a three term recursion:

\[
a = \langle x \cdot bf_k(x), bf_k(x) \rangle / \langle bf_k(x), bf_k(x) \rangle
\]

\[
b = \langle bf_k(x), bf_k(x) \rangle / \langle bf_{k-1}(x), bf_{k-1}(x) \rangle
\]

\[
bf_{k+1}(x) = (x-a) \cdot bf_k(x) - b \cdot bf_{k-1}(x).
\]

The DPD space \( S \) is composed of 24 orthogonal basis functions, 6 for the main and 3 memory functions (taps). Fig. 4.32 presents the PA output spectrum with indirect learning DPD architecture by using adaptive selection of basis functions. The black curve denotes the PA output before start of DPD. The blue curves present the spectrum of PA output at each iteration after running the DPD adaptations.

![PA Output Spectrum](image)

**Figure 4.32:** PA output spectrum of indirect learning DPD of using adaptive selection of basis functions.
Unlike in a conventional DPD system that the PA output spectrum would be already suppressed to noise level at second or third DPD adaptation, the DPD by using the proposed method needs a few adaptations more to achieve the maximal compensation. The blue curves in Fig. 4.32 present the transition from initial state to optimal compensation. The two threshold values $\delta_1$ and $\delta_2$ are dependent on the current and historical average value of $\gamma_{k,q}$, as defined by:

$$\delta_1 = 2 \cdot \gamma_{\text{ave}}^{m}$$

(4.97a)

and

$$\delta_2 = 0.2 \cdot \gamma_{\text{ave}}^{m},$$

(4.97b)

where

$$\gamma_{\text{ave}}^{m} = 0.5 \cdot \gamma_{\text{ave}}^{m-1} + 0.5 \cdot \sum_{k=0}^{K-1} \sum_{q=1}^{Q} \frac{\gamma_{k,q}}{K \cdot Q}.$$  

(4.98)

Since the threshold values $\delta_1$ and $\delta_2$ are not optimized, the convergence speed of the approach is relatively low. The red curves denote the spectrum of PA output at the last iterations and the ACP disappears under the noise level. It is obvious that the DPD achieves a very good linearization performance in conjunction with this proposed method of adaptive selection of basis functions.

In Fig. 4.33 the number of basis functions contained in DPD space $S$ is plotted against DPD adaptation number. Instead of using all 24 basis functions, only the basis functions with high projection from the DPD residual are involved to calculate the coefficient updates. The initial DPD space $S$ has only 2 basis functions. At run time the number of basis functions dynamically follows the residual of DPD system. At first the number of involved basis functions is increased, because the residual of DPD is relatively large and more basis functions are needed to reduce the DPD residual. After the maximum of 8 basis functions is reached, the residual decreases significantly. Consequently, the residual projection on each basis function becomes negligible small and the basis functions are removed from the DPD.
space $S$ step by step. At the last adaptations the number of involved basis functions levels out around 3. The number of basis functions involved in DPD space $S$ is dramatically reduced from 24 to 3 in the steady state.

If all available basis functions have been used, the condition number of the matrix $M_y$ with all 24 orthogonal basis functions for single-carrier UMTS signal has a value over 1200. In contrast to this, the condition number of the matrix $M_{bf}$ only consisting of the selected basis functions is much smaller. In most cases the condition number is less than 40, as depicted in Fig. 4.34. With the reduced basis functions as well as the condition number, the DPD performance will not be deteriorated. Instead the DPD performs even more stable, because less basis functions need to be taken into account. Thus, complexity and
computational cost of this method is reduced significantly as well as the robustness of the DPD is improved.

Naturally the convergence speed can be improved, if one increases the number of initial basis functions or optimizes the threshold $\delta_1$ and $\delta_2$. In the next simulation only the threshold $\delta_1$ is changed to be two-third of its value in equation (4.97a). With this relaxed threshold more basis functions with relative low correlation with the DPD residual can be added to the basis function at an early phase than in the previous simulation.

In Fig. 4.35 one can observe that the number of blue curves between the black and red curves is much lower than the number of blue curves in Fig. 4.32, i.e. the DPD needs less iterations to reach its maximal performance in terms of ACP suppression.

![PA Output Spectrum](image)

**Figure 4.35: PA output spectrum of indirect learning DPD with relaxed threshold.**

Fig. 4.36 presents the number of basis functions at each DPD adaptation. The number of involved basis functions increases rapidly to 11 for the relaxed threshold $\delta_1$. Moreover, the number of basis functions drops also rapidly back to a low level with decreased DPD residual. In practice, the optimal coefficients can be loaded by start of DPD to achieve maximal performance. At the same time one can set the DPD space $S$ be an empty set. If the DPD residual is small enough, one does not even need to adapt the DPD coefficients any more. Because the DPD residual is so small, no basis functions need to be added in
DPD space $S$. In the other case with certain measurable DPD residual, according to DPD performance requirements, the basis functions would be added to the DPD space $S$.

![Adaptive Selection of Basis Functions](image1)

Figure 4.36: Number of basis functions of indirect learning DPD with relaxed threshold.

In addition the proposed method for the DPD with direct learning architecture is simulated as well. The equations (4.90) and (4.93) are applied to verify the idea of saving computational costs by replacing the feedback signal $y[n]$ by the original signal $x[n]$. The DPD works fine with these two equations, because the nonlinearities of PA model are completely compensated, as shown in Fig. 4.37.

![PA Output Spectrum](image2)

Figure 4.37: PA output spectrum of direct learning DPD of using adaptive selection of basis functions.
The number of basis functions and the condition number for each adaptation are presented in Fig. 4.38 and Fig. 4.39, respectively. Similar performance is observed for both the indirect and direct learning DPD structures. The number of basis functions used in coefficients updates computation and the condition number of the system of equations are both much smaller than these values for conventional method. Furthermore, the direct learning DPD exhibits much lower computational cost than indirect learning DPD, which would be investigated in Chapter 5.

![Figure 4.38: Number of basis functions used in DPD adaptation of direct learning architecture.](image)

![Figure 4.39: Condition number of direct learning DPD using adaptive selection of basis functions.](image)

4.5.4 Experimental Results

To verify the effectiveness of the DPD using adaptive selection of basis functions, an offline DPD system with memory compensation is implemented in FPGA. The computation of
polynomial coefficients is performed offline using Matlab on a PC. The DPD with separate LUTs structure is implemented in signal path in FPGA to modify the original input signal, shown in Fig. 3.4. The PA used in the experiments is a Class-AB power amplifier with average output power of 50W.

In the lab experiments the direct learning architecture is used to estimate the DPD coefficients. This method with adaptive selection of basis functions performs very well in conjunction with the orthogonal polynomial DPD in terms of suppression of ACP and reduction of EVM. In Fig. 4.40, the spectrum of PA output for a 4-carrier UMTS signal is depicted without and with the described DPD. The ACPs are suppressed over 20 dB through the DPD in conjunction with adaptive selection of basis functions.

![Figure 4.40: PA output spectrum of direct learning DPD of using adaptive selection of basis functions.](image)

In Fig. 4.41 the number of basis functions for each adaptation is presented. Instead of using all 24 basis functions, only the basis functions with large correlations to the residual of DPD system are introduced to estimate the updates of DPD coefficient. At run time, the number of basis functions increases to the maximum 9 and then decreases and levels out at 4. The number of basis functions involved in DPD space $S$ is much less than 24, which is an absolute must for conventional DPD not using the method of adaptive selection of basis functions. With reduced number of basis functions the DPD performs at least as good as conventional DPD.
Figure 4.41: Number of basis functions used in DPD adaptation of direct learning architecture.

The condition numbers at each DPD adaptation are very small, below 20 in most of the time, as shown in Fig. 4.42, which indicates a very stable and reliable DPD system. The simulation and lab experiments quite well match with each other and demonstrate that this method is very effective in reducing the nonlinearities of PAs and improving the stability of the DPD system. The method of adaptive selection of basis functions can be used for most basis functions based DPD algorithms, including the Volterra series.

Figure 4.42: Condition number of direct learning DPD of using adaptive selection of basis functions.

The approach of using adaptive basis functions provides two essential advantages. First the stability is significantly improved by reducing the number of basis functions used in computation of DPD updates. As a consequence, the DPD system of equations composed of these selected basis functions exhibits a much smaller condition number than the
conventional DPD system. The stability of the entire DPD system is therefore significantly improved. The other advantage is the saving of computational cost. The computational complexity is heavily reduced, because the number of equations is strongly reduced, which improves the execution time of each iteration. Furthermore, in direct learning DPD architecture one does not need to calculate the basis functions $bf_{q,k}(y[n])$ of the feedback signal $y[n]$ and instead, one can use the basis functions $bf_{q,k}(x[n])$ of the original input signal $x[n]$, which have already been computed in PD. Thus, the computational cost can be significantly reduced furthermore. The detailed investigation of the computational cost can be found in Chapter 5.

4.6 Combination of Adaptive Selection of Basis Functions and Iterative Methods

The presented method of dynamical selection of basis functions and the iterative methods are two very effective methods to improve the stability and reduce the computational cost of a DPD system. Because the iterative methods have no special requirements on the system of equations, the iterative methods can be used to solve the system of equations based on the selected basis functions with some adjustments. The equations (4.87) and (4.91) would be multiplied with the Hermitian transpose (conjugate transpose) of matrix $M_{bf}$. Since the effects of old DPD coefficients have already implicitly been taken into account in the equations (4.87) and (4.91), i.e. only the updates of DPD coefficients are calculated, the initial guess of iterative methods should be set to zero instead of the old DPD coefficients.

4.7 Fast Load Changing of Power Amplifiers

Another critical issue of offline DPD formulated as a research task is how to deal with fast load changes. This problem is rarely discussed neither in the subject of PA modelling nor in the subject of DPD. In PA modelling a mathematic model is used to describe the stationary PA behaviour assuming that the PA behaviour does not change depending on the input signal. It is very likely that the estimated PA model only achieves an acceptable accuracy for the training signal used for tracking of the PA behaviour. The accuracy of the PA modelling
may deteriorate for other signals with different bandwidth and different power density functions, because a certain test signal can not stimulate all dynamical effects of a PA. It is also observed in practice that DPD exhibits different characteristics with respect to different mobile communication standards and signal bandwidth.

In most applications only the specifications for full-load case are defined by the governments or international telecommunication associations, e.g. the test-signal, the ACP and EVM requirements. In some applications the dynamical rescheduling of available resources results in significant change of the signal dynamic range, also known as load change. For instance, the HSDPA (High Speed Downlink Packet Access) for WCDMA (Wideband Code Division Multiple Access) has a short Transmission Time Interval (TTI) of 2ms. The HS-SDCH (High Speed Downlink Shared Channel) is mapped to the shared code resource composed of up to 15 codes, depending on system operator settings and system capacity etc., see [67]. A possible HS-DSCH code and time structure is presented in Fig. 4.43. This section discusses the consequences of fast load changes, namely the fast change of the input signal dynamical range, for the offline DPD and its solution.

![Figure 4.43: HS-DSCH code and time structures.](image)

### 4.7.1 Problem Analysis

The problem of fast load change can be specified in two scenarios, i.e. switching from full load to low load and the opposite case, switching from low load to full load. The first scenario, switching from full load to low load, does not raise any problem for DPD system. In the full load operating model, the samples distributed over the entire PA input dynamic
range are used to calculate the DPD coefficients and thus the DPD can accurately estimate the entire inverse characteristic of the PA. If in the next 2ms only the samples of low power (low load case) are transmitted, the DPD can still work properly with current coefficients. However, in second scenario only the samples of low power level have been transmitted for a while and hence the DPD only has these samples of low power available for estimation of the entire PA inverse characteristic. Since the information of the PA characteristic for high input power level is missing, the DPD is incapable of estimating the inverse PA gain in the range of high input power level. It is very likely, that the PA inverse characteristic at high input power level would be incorrectly estimated. Consequently, if full load case is coming in next millisecond, the DPD can not provide correct predistortion gain for these samples of high power level. Furthermore, the update rate of offline DPD is relative low, ranging from hundreds of milliseconds to seconds. All these result in spectrum regrowth namely interference into neighbour spectrum or violation of the standard specifications, when the second described scenario takes place.

### 4.7.2 Method of Attaching Old Samples

There are some smart DPD algorithms, such as Piecewise Polynomial Approach (PIPOP), see [68], which divides the entire PA input dynamic range into several small intervals. The DPD coefficients for each interval are relatively independent of these coefficients for other intervals and hence can be updated separately depending on the power level of the available samples. In different load case the DPD only updates its characteristics in the power range, in which the samples are transmitted. In this way the problem of fast load changes described in the last paragraph can be solved without taking additional measures. But generally such DPD algorithms suffer from their own disadvantages, such as too many coefficients and difficult to be orthogonalized. In contrary, the widely used polynomial and orthogonal polynomial algorithms have a set of coefficients for entire PA input power range and each coefficient directly influences entire range of DPD characteristic. Therefore, one has to take some extra measures to handle the problem raised by fast load changes. To be more precise, one needs to deal with the problem of switching from low load to full load. One possible
solution applied in some applications is the approach of Multi-DPD, as shown in Fig. 4.44. According to the current load case one of the PDs is switched on by the control unit to provide the proper correction for the PA. Furthermore, a similar control unit is needed to select one of the PDs by updating of its coefficients, which is not depicted in Fig. 4.44. One disadvantage of this approach is the increased latency in the transmitter, because the control unit needs to estimate the load case over a certain time slot. Moreover, the permanent switching between these PDs also introduces spectrum fluctuation and additional instable factor in the entire system. Thus, a few dB margin for spectrum regrowth due to the switching between these PDs is still required in this approach.

![Multi-DPD approach for dealing with fast load change.](image)

Since the problem of fast load change is identified as a lack of information in certain PA power range, the solution could be providing the missing information in that PA power range. The intention of this idea is to add old samples with high power level to the current captured feedforward and feedback samples, as shown in Fig. 4.45.

![Method of attaching old samples of missing power level to new captured samples.](image)
Because all these new captured samples in the blue box cover only the lower half of PA input power range, the PA inverse characteristics in the other half of the input power range can not be estimated correctly without the samples of the upper half range. Fortunately, the samples of high input power captured at earlier DPD adaptations contain the desired information of the PA behaviour in the high power range, as shown in the two red boxes in Fig. 4.45. Therefore, all these samples in blue and red boxes can be used to calculate the DPD coefficients for calculation of the new DPD coefficients.

4.7.3 Experimental Results

The simulations are not carried out for the method of attaching old samples, because all available PA models are stationary, which do not change their behaviour depending on various load cases. As a consequence, one would not observe any difference, whether the attached samples are old or not. One would get exactly the same DPD coefficients for the two cases of attaching old samples and feeding these samples through PA. In practice, it is expected that the PA behaviour changes slightly for different load cases because of the temperature drifting and/or other effects. This change of PA behaviour would influence the performance of the method of attaching old samples, which can only be observe in hardware experiments instead of simulations. It is worthwhile to point out that the method of attaching old samples works only well, if the PA characteristics do not change a lot with the temperature drifting. As usual an offline DPD with separate LUTs structure is implemented in the signal path in FPGA, as shown in Fig. 3.4. The PA used in the experiments is a Class-AB power amplifier with average output power of 50W.

The software routine for the proposed method is presented in Fig. 4.46. Before the calculation of DPD coefficients the maximal power of the feedback signal would be analysed to evaluate the current load situation. In case of low load the old samples with high power would be attached to the captured new samples. Subsequently the DPD coefficients would be calculated and downloaded to FPGA. If full load occurs, the new samples with various high power levels would be saved to replace the old saved samples and in this way the latest PA behaviour can be stored for later use.
The UMTS test model signal is modified to have short 2ms Transmission Time Interval (TTI) of different loads. As shown in Fig. 4.47, the 10ms UMTS frame is composed of three pieces of full load and two pieces of 25% load (6dB less power). The output power of the test PA is about 47dBm for the full load and 41dBm for the 25% load.

Figure 4.47: I/Q test signal with 2ms Transmission Time Interval of full load and 25% load.
For the first test, the captured feedforward samples and feedback samples contain both the samples of full load and 25% load, as presented in Fig. 4.48. The red curve and blue curve, ranging from 0 to 1, denote the amplitude of the feedforward signal and feedback signal, respectively. It is the best case, because all information on the entire PA input power range is available for estimation of the DPD coefficients.

![Figure 4.48: Amplitude of captured samples of full load and 25% load.](image)

Fig. 4.49 depicts the spectra of PA output of dynamic load with and without DPD. As expected, the DPD performs as good as in case of full load, where the blue curve with DPD exhibits over 20dB additional suppression of ACP.

![Figure 4.49: PA output spectra with signal burst of difference load cases (I).](image)

It is worthwhile to point out that the spectra in Fig. 4.49 and some of the following figures are actually the area shaped by the PA output spectra of constant output power at 47dBm highlighted by red curve and 41dBm highlighted by yellow curve. Within the signal
bandwidth one can observe that the difference between the upper edge (red marked) and lower edge (yellow marked) is exactly 6dB.

It is very likely that only samples of low power level are transmitted in the time slot of data capture. Thereby, the DPD only has the low power samples available to compute the inverse characteristics of entire PA power range. To verify this case in lab experiments, only the samples of 25% load (low load) are used to calculate the DPD coefficients. Furthermore, no additional measures are taken to handle the lack of information in PA behaviour for high power level.

It can be observed in Fig. 4.50 that the lower edge of the blue curve (with DPD) exhibits similar performance in terms of ACP suppression in comparison to the lower edge of the blue curve shown in Fig. 4.49. Thus one can state that the PA nonlinearities for 25% load case are sufficiently compensated. But in contrary, the upper edge of the full load has only 2 to 4 dB improvement on the ACP value, which indicates a failure of DPD for the high power input signal. In worst case the upper edge of blue curve would be even higher than the upper edge of the black curve (without DPD).

As the problem of fast load change has been proven in lab experiments, the method of attaching old samples with high input power level is used to overcome this problem. In order not to increase the computational cost, only dozens of old samples of high power are attached to the new captured samples as shown in Fig. 4.51.
This time the upper edge of the blue curves for full load has been suppressed as well, presented in Fig. 4.52. The green curve covered by the blue one presents the PA output spectrum for the case, that a large number of samples of high power level are available for DPD adaptation. The blue curve exhibits slightly worse performance than the green one. Since one does not want to spend too much memory and increase too much computational cost by attaching too many old samples, the amount of attached old samples is much smaller than the amount of new captured samples.

In order to achieve better DPD performance and avoid attaching too many old samples, one can allocate a higher weighting factor to the old samples than the new ones to reduce the
number of attached old samples. In the further experiments the system of equations is multiplied by the weighting function $f_w$, as given by:

$$x_{DPD} \cdot f_w = M_y \cdot A \cdot f_w,$$

(4.99)

In the following example the attached samples weight twice as the captured new samples and the weighting function is defined by:

$$f_w = \left[ f(x_{\text{new}}), \ldots, f(x_{\text{old}}), \ldots \right]^T,$$

(4.100)

with $f(x_{\text{new}}) = 1, f(x_{\text{old}}) = 2$.

Under the same conditions the blue curve with weighting factor has slightly better performance in terms of ACP suppression than the green one without weighting factor, as illustrated in Fig. 4.53.

![Figure 4.53: PA output spectra using method of attaching old samples and weighting factor for attached samples.](image)

Fig. 4.54 presents the DPD characteristics for the described experiments. The black curves denote the DPD characteristics adapted by captured samples of full load, which provides the best DPD performance in terms of ACP suppression. The blue and green curves present the DPD characteristics by using the method of attaching old samples with and without weighting factors. It is obvious that the blue curves with weighting factor of 2 are located much closer to the black curves in comparison to the green curves. Therefore, the DPD of
blue curves with weighting factor performs slightly better than the DPD of green ones without weighting factor.

![Figure 4.54: Estimated DPD characteristics with and without weighting factor.](image)

Furthermore, to investigate the temperature drifting effect caused by various load cases, the following experiments are carried out. At first, the PA was operated with full load for a while, and then the source signal would be switched to 25% load. After the PA cooled down a few minutes, the source signal is switched back to full load again. The Fig. 4.55 and Fig. 4.56 illustrate the two cases without and with attaching of old samples, respectively.

![Figure 4.55: PA output spectra without attaching old samples.](image)

In both figures the black and green curves present the PA output spectrum of constant output power 47dBm and 41dBm with optimal DPD coefficients. The blue curve describes the PA output spectrum after the full load signal is switched on again. For the case that the DPD is adapted without the information in PA behaviour for high power range, i.e. without
attaching of old samples, the ACP values increases dramatically and largely exceed the permitted value from specifications. The spectrum regrowth is even higher than the original PA output spectrum without DPD, which is not depicted in Fig. 4.55.

![PA output spectra with attaching of old samples.](image)

For the case of attaching old samples the blue curve is only 1.5 to 2 dB slightly higher in comparison to green curve of optimal DPD coefficients, as shown in Fig. 4.56. Because the PA has cooled down after sending low power signal (25% load) for a while, the PA behaviour information contained in old samples does not match perfectly the current PA behaviour. But the deviation from the optimal performance is very small, which still proves the effectiveness of the proposed method of attaching old samples. One can draw the conclusion, that the new method of attaching old samples can handle the problem of fast load change. This method ensures that the missing information of PA behaviour is refilled by the old samples and needs only extra handling in DPD adaptation but does not introduce additional latency in the signal path. Moreover, the modification of existing DPD adaptation process is so small, that it can be easily done as a software routine.

### 4.8 ET Temporal Misalignment Analysis and its Compensation Methods

DPD technique can not only be used in conventional PA concepts such as Class-AB, Class-B PAs, but also in other PA techniques. Conventional PAs with constant power supply have only high efficiency at high output power. In other words, conventional PAs only work effectively in saturation region. However, PAs are seldom operated in saturation region,
because the proportion of the high power signal is very low in the signal of Gauss distribution and hence PAs are usually operated with a huge back-off for the input signal of dynamic amplitude. The fundamental idea of Envelope-Tracking (ET) technique is to dynamically adapt the PA power supply voltage to the envelope of the input RF signal by a highly efficient supply voltage modulator [69], [70]. Thus, PAs in ET system are always operated close to their saturation region, i.e. in high efficiency region. In practice, a DPD technique cannot sufficiently compensate the nonlinearities generated in an ET system, because some new imperfections and interference in ET system cannot be described by conventional PA models and hence cannot be compensated by DPD.

One of the major concerns of ET systems is the temporal mismatch between the RF signal and its envelope signal, as shown in Fig. 4.57. Since the two input signals of the ET system, the RF signal $S_{RF}(t)$ and its ET signal $S_{ET}(t)$, are delivered by two different signal paths, these two signals might arrive asynchronously at the PA inputs. The resulting temporal difference is denoted by $\tau$ and can be either positive or negative. This temporal mismatch between RF signal and ET signal can deteriorate the EVM and ACP value. Furthermore, the temporal mismatch might disturb other compensation processes in the transmitter, e.g. DPD and IQ imbalance compensation.

The delays in the three paths depicted in Fig. 4.57, such as ET path, feedforward path and feedback path can be accumulated and presented by $\tau_{ET}$, $\tau_{FW}$ and $\tau_{FB}$, respectively. In practice none of the three delays is known and without loss of generality a delay module can

![Figure 4.57: ET system.](image-url)
be placed in the ET path to compensate the temporal misalignment of RF signal and its ET signal. In simulations the compensation delay $\tau_{\text{com}}$ can be positive and negative depending on the difference between $\tau_{\text{ET}}$ and $\tau_{\text{FW}}$. Because it is difficult to implement a negative delay in real hardware, a delay offset is introduced in the feedforward path in such a way that the delay $\tau_{\text{ET}}$ is always smaller than the delay $\tau_{\text{FW}}$. In this way the delay mismatch can be compensated by a positive delay $\tau_{\text{com}}$ in ET path.

The problem of compensating the delay mismatch between the RF signal and its ET signal has been investigated in [71], [72]. The existing solutions for delay compensation are based on the assumption, that the PA system is completely linearized by using DPD. But in reality the DPD performance is strongly impacted by an uncompensated temporal misalignment between RF signal and its ET signal. Therefore, it is not reasonable to assume, that the PA is perfectly linearized before the compensation of the temporal misalignment. This work presents two novel approaches to analyze and compensate the temporal misalignment without the assumption of a linearized PA system. By contrast, in the new approaches the nonlinearities of PAs are used to estimate the sign of the time difference between $(\tau_{\text{ET}} + \tau_{\text{com}})$ and $\tau_{\text{FW}}$ instead of being neglected. Consequently, one can adjust $\tau_{\text{com}}$ in the ET path step by step to achieve perfect match between RF signal and ET signal.

#### 4.8.1 Analysis in Time Domain

To demonstrate the impact of a temporal mismatch between the RF signal and its ET signal, a periodic signal with symmetrical amplitude (red marked) is supposed to be fed to an ET system, as shown in Fig. 4.58. The black curve denotes the desired ET signal of the RF signal. At top left the normalized RF signal arrives after the ET signal at the PA. At top right the reverse case is illustrated, i.e. the ET signal reaches the PA before the RF signal. Under the assumption of a perfectly linearized PA due to DPD the PA output signal is approximately determined by the minimal value of the RF input signal and the ET signal. Therefore, the PA output signal described by blue curves on the bottom of Fig. 4.58 is
exactly symmetrical and one can not extract any useful information from the two exactly identical blue curves.

If the PA nonlinearity is not completely compensated by DPD, one would observe a totally different PA output signal presented by the green curves at the bottom of Fig. 4.58. At the left side the rising ramp of ET signal in black comes earlier than the RF signal, which means that the PA supply power is higher than the expected value. The PA would therefore provide more gain than desired value. Otherwise, for the falling ramp the PA exhibits less gain than expected, since the PA power supply is smaller than the desired value. At the right side the green curve looks exactly like the green curve at the left side in the mirror. In the manner, one can integrate the area under the green curves (amplitude of the feedback signal) for the rising and falling ramps to extract the information, if the RF signal arrives at the PA before or after the its ET signal. Thereby, one is able to derive the sign of the time difference between $(\tau_{ET} + \tau_{COM})$ and $\tau_{FW}$ from the integrated area below the rising and falling ramps. Subsequently, one can increase or decrease the delay $\tau_{COM}$ to minimize the temporal mismatch step by step. Otherwise, if the areas under the two parts of the green curves are as large as each other, one can state that the temporal misalignment between the RF and its ET signal is compensated perfectly.

![Figure 4.58: Symmetrical signal amplified by ET system.](image)
Simulation Results

The well-known Cann’s model [72], [73] is often used for the simulation of the ET system, which has a second input of ET signal to describe the PA behaviour under dynamic power supply. The used AM-AM distortion of a Cann’s model with its technology dependent parameters described in [72] is defined by:

$$ RF_{out} = \frac{g \cdot RF_{in}}{1 + \left( \frac{g}{L} |RF_{in}| \right)^{\frac{1}{s}}} $$  \hspace{1cm} (4.101)

with \( g = 1.24 + 0.38 \cdot ET \), \( L = 0.09 + 1.15 \cdot ET \) and \( s = 5 \).

Fig. 4.59 presents the simulation results for the two cases, that the RF signal has a negative or positive delay with respect to its ET signal. The RF signal has a delay of 8 clock cycles and the ET signal has the delay of 12 clock cycles for case (a) and 4 clock cycles for case (b). A symmetrical triangle wave as test signal is fed to Cann’s model. The areas under the rising ramps and falling ramps would be averaged over hundreds of single triangles. After several adaptations the ET delay would be reduced or increased step by step and hence match the RF delay, as shown at the bottom of Fig. 4.59.

In further simulations a fractional delay is introduced in the feedforward path to verify the effectiveness of this method. One integer delay is split into 128 fractional delays, which are implemented by 128 fractional delay filters. In Fig. 4.60 the calculated areas under the ramps
and the estimated fractional delays are presented, respectively. Through comparing the areas under the rising ramp and falling ramp of training signal the estimated fractional delay approaches the set value of fractional delay 25 (25/128ths of a sampling clock cycle) and levels out at this set value. The slight deviations presented in the figure on the bottom of Fig.4.60 are caused by the environmental noise (AWGN) added to the PA output during the simulations.

![Figure 4.60: Compensation of temporal mismatch with fractional delay.](image)

One of the difficulties in this method is to exactly find the peak point of the triangle signal. The test signal with sharp peaks usually has a very broad bandwidth, which can not pass through the transmitter without lost of some high frequency parts. Furthermore, the peaks of the feedback signal may be located between the samples. Thereby, interpolation technique can be used to overcome this problem.

### 4.8.2 Analysis in Frequency Domain

Generally symmetrical signal is not available during the operating time, because one does not transmit symmetrical signals in practice. But one can still use this method by comparing the ratio of the areas under the rising and falling ramps for both PA input signal and PA output signal to extract the sign of the temporal misalignment. However, this modified method not only introduces much more computational cost but may deteriorate the accuracy. Therefore,
further investigation of temporal misalignment is made in frequency domain. An ET system is a system with two inputs and one output, as described by:

\[ y(t) = f_{PA}(f_{RF}(x(t)), f_{ET}(x(t))) \Rightarrow y(t) = f(x(t)), \]  

(4.102)

where \( x(t) \) and \( y(t) \) denotes the RF input and output signal of the ET system. The functions \( f_{RF}(x) \) and \( f_{ET}(x) \) present the transfer functions of feedforward path and ET path, respectively. The equation (4.102) describes the case of no temporal mismatch between the ET path and RF path. Otherwise, if there is any temporal mismatch between these two paths, the equation (4.102) can be changed to:

\[ y(t) = f_{PA}(f_{RF}(x(t)), f_{ET}(x(t + \tau))) \Rightarrow y(t) = f(x(t), x(t + \tau)), \]  

(4.103)

where the delay \( \tau \) can be positive or negative. It is obvious that the PA output \( y(t) \) not only depends on the current input \( x(t) \) but also on previous or future input signal \( x(t + \tau) \), depending on the sign of delay \( \tau \). Thus the memoryless PA model starts to have memory effect due to the temporal misalignment of the RF signal and its ET signal, as shown in equation (4.103).

Figure 4.61: AM-PM distortion of a PA based on a GaN transistor.

In this work, the PA used in the ET system is based on a GaN transistor, which exhibits extremely low and flat phase distortion with varies drain voltages, as shown in Fig. 4.61. One can observe that the phase distortion in the entire PA operating range is smaller than 5 degree tested at frequencies 2110MHz, 2140MHz and 2170MHz and denoted by blue, red
and green curves, respectively. Moreover, in paper [74] the authors notice similarly low phase distortion and relatively low scatter of the AM-AM and AM-PM spreads for the PA based on GaN technology, which indicates very weak memory effects. For this reason, it is also reasonable to neglect the memory effects of the PA based on GaN technology in this work. Therefore, a novel PA behaviour model for ET system, as illustrated in Fig. 4.62, is developed regarding the aspect of extremely low phase distortion and extremely weak memory effects. The basic idea in this PA model is that the dynamic power supply affects only the output power of the PA and has almost no influence on the PA phase distortion, i.e. the temporal misalignment of ET and RF signal causes only a between in AM-AM distortion and AM-PM distortion.

![Figure 4.62: Novel PA behaviour model for ET system.](image)

The intention of this PA model is to find the relationship between the temporal misalignment and PA output spectrum. The PA model is further simplified to have only third order AM-AM distortion, since the attention is focused on IMD3. To ease the derivation the PA baseband model is defined by:

\[
y[n] = x[n] \cdot \left( a_1 + a_3 \cdot |x[n]|^2 \right) \cdot e^{j \phi(|x[n]|)} ,
\]

where \(a_1\) and \(a_2\) are two real-valued coefficients. The phase distortion, see [75], can be expressed as:

\[
\phi(|x[n]|) = 2k \cdot |x[n]|^2 .
\]
The AM-PM distortion has a parabolic form. In Fig. 4.63 the assumed AM-PM distortion with positive and negative coefficient $k$ is presented.

As the temporal mismatch only results in the temporal shift of AM-AM distortion [75], a variable $\tau_{AM}$ is introduced to present the delay effect caused by the temporal mismatch between the RF signal and its ET signal, as given by:

$$y[n] = \cos(\omega n) \cdot \left( a_1 + a_3 \cdot \cos^2(\omega n + \tau_{AM}) \right) \cdot e^{j \cdot k \cdot \cos^2(\omega n)},$$  \hspace{1cm} (4.106)

where the two-tone baseband test signal $\cos(\omega n)$ is used again as a stimulus signal in further discussions. In [75] the delay between the AM-AM and AM-PM distortions is introduced to explain the asymmetrical IMDs of conventional PAs in RF band. In this work a similar effect is supposed to be raised by temporal mismatch between the RF signal and its ET signal. Moreover, a clear relationship between the delay $\tau_{AM}$ and PA output spectrum is derived from the PA baseband model. The detailed derivation can be found in Appendix C. One important interim result is described by:

$$y'[n] = \left( a_1 + \frac{a_3}{2} \right) \cdot \cos(2 \omega n + 2 \tau_{AM}) \cdot \left( 1 - k \cdot \cos(2 \omega n) + j \cdot k \cdot \cos(2 \omega n) \right) \cdot \cos(\omega n).$$  \hspace{1cm} (4.107)

After some reasonable simplifications the IMD3 of interest is given by:
\[ y'_{3,ul}[n] = e^{j3\alpha n} \frac{1}{2} \left[ \frac{a_1}{4} \cos(2\tau_{AM}) - \frac{a_3}{8} \cos(2\tau_{AM}) - \left( a_1 + \frac{a_3}{2} \right) \frac{k}{2} \right] + j \left[ \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} + \frac{a_3}{8} \cdot \cos(2\tau_{AM}) + \frac{a_3}{4} \sin(2\tau_{AM}) \right] \]

One can observe a difference in the imaginary parts between the upper and lower IMD3. Because only the power of IMD3 is of interest and the real parts for both IMD3 are identical, one merely needs to compare the absolute value of these imaginary parts to find out which’s power is larger than which’s. If a relationship between the asymmetry of IMD3 and the time delay \( \tau_{AM} \) can be found, one is able to make a statement of the delay mismatch between the feedforward and ET paths regarding the PA output spectrum. The imaginary parts of lower and upper IMD3 are given by:

\[ IMD3_{L_{im}} = \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} \cdot \frac{a_3}{4} \sin(2\tau_{AM}) \]

and

\[ IMD3_{U_{im}} = \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} \cdot \frac{a_3}{4} \sin(2\tau_{AM}) \]

For conventional PAs the coefficients \( a_1 \) and \( a_3 \) are positive and negative, respectively. But in ET systems the sign of \( a_3 \) strongly depends on the ET signal and hence is not always negative. Therefore, one has three variables \( a_3 \), \( k \) and \( \tau_{AM} \) with possible positive and negative sign to derive the relationship between the IMD3 asymmetry and temporal
mismatch $\tau_{AM}$. Eight cases for all available combinations of these three variables need to be analysed, as presented in the Table 4. The term $A$ and term $B$ are defined by:

$$A = \left( a_1 + \frac{a_3}{2} + \frac{a_3k}{4} \cdot \cos(2\tau_{AM}) \right) \cdot \frac{k}{2}$$

(4.111)

and

$$B = \frac{a_3}{4} \sin(2\tau_{AM}),$$

(4.112)

respectively. Furthermore, the delay $\tau_{AM}$ is assumed to be very small and hence has the same sign as $\sin(2\tau_{AM})$.

<table>
<thead>
<tr>
<th>$a_3$</th>
<th>$k$</th>
<th>$\tau_{AM}$</th>
<th>Term $A$</th>
<th>Term $B$</th>
<th>$IMD3 _U$ vs. $IMD3 _L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+ (decrease $\tau_{COM}$)</td>
<td>+</td>
<td>+</td>
<td>&gt;</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>- (increase $\tau_{COM}$)</td>
<td>+</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>+ (decrease $\tau_{COM}$)</td>
<td>-</td>
<td>+</td>
<td>&lt;</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>- (increase $\tau_{COM}$)</td>
<td>-</td>
<td>-</td>
<td>&gt;</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>+ (decrease $\tau_{COM}$)</td>
<td>+</td>
<td>-</td>
<td>&lt;</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>- (increase $\tau_{COM}$)</td>
<td>+</td>
<td>+</td>
<td>&gt;</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>+ (decrease $\tau_{COM}$)</td>
<td>-</td>
<td>-</td>
<td>&gt;</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>- (increase $\tau_{COM}$)</td>
<td>-</td>
<td>+</td>
<td>&lt;</td>
</tr>
</tbody>
</table>

Table: 4 Relationship between IMD3 asymmetry and temporal mismatch in ET system.

The bracket $a_1 + \frac{a_3}{2} + \frac{a_3k}{4} \cdot \cos(2\tau_{AM})$ in the term $A$ is dominated by the coefficient $a_1$, because the coefficient $a_1$ is much larger in comparison to $a_3$. Therefore, the sign of the term $A$ in equation (4.111) is determined by the coefficient $k$. If the term $A$ and term $B$ have the same sign, the power of $IMD3 _U$ is larger than the power of $IMD3 _L$. For the opposite case the power of $IMD3 _L$ is larger the power of $IMD3 _U$.
If the signs of PA coefficients $a_3$ and $k$ are known, through the information of IMD3 asymmetry the sign of time delay $\tau_{AM}$ can be read from the Table 4. For instance, in the first row the term $A$ is positive, if the power of $IMD3_U$ is larger than the power of $IMD3_L$, the term $B$ should have the same sign as the term $A$, i.e. positive. Therefore, the term $\sin(2\tau_{AM})$ must be positive and the time delay $\tau_{AM}$ is larger than zero, which means that the RF signal reaches the PA earlier than its ET signal. The compensation delay $\tau_{COM}$ in ET path should be decreased in this context. In this way, the delay $\tau_{COM}$ can be adapted step by step according to the sign of $\tau_{AM}$ until the IMD3 becomes symmetrical.

**Simulation Results**

Simulations are performed in Matlab to verify the derived relationship between IMD3 asymmetry and the ET temporal misalignment. In simulations only the AM-AM distortion in Fig. 4.62 is replaced by the Cann’s model described in equation (4.101) and a single-carrier UMTS signal is fed to the ET system model instead of a Two-Tone test signal. The $k$ value of AM-PM distortion is set to be positive. The coefficient $a_3$ is proven to be positive, if 3rd polynomial is used to approximate the AM-AM distortion modelled by the Cann’s model. One just monotonously sweeps the delay $\tau_{COM}$ to see if the IMD3 asymmetry changes according to the rule, illustrated in Table 4.

Three spectra of PA output with different temporal mismatch value from negative to positive are presented in Fig. 4.64. The figure in the middle without any temporal mismatch exhibits almost symmetrical ACP. The Fig. 4.64a at left side has a higher value at lower ACP than the value at upper ACP, which confirms to the negative sign of $\tau_{AM}$ shown in second row in Table 4. The Fig. 4.64c at the right side exactly presents the opposite case that the ET path has larger delay than the feedforward signal path.
In Fig. 4.65 the ACP asymmetry is plotted over the temporal misalignment from -4 system cycles to +4 system cycles. It is obviously clear that the asymmetry of the ACP changes monotonously with increased time delay between the RF path and ET signal path. In case of a perfect compensation upper ACP and lower ACP are almost symmetrical. This method of signal analysis in frequency domain does not require any special training signal, so that it can be applied during the PA operating time.

![Figure 4.65: ACP Asymmetry over temporal misalignment between RF signal and its ET signal.](image)

4.8.3 Experimental Results

In lab experiments the test platform for DPD is extended by an ET signal path composed of a 12bit DAC, low pass filter and an ET modulator. The PA used in the ET system is based on GaN technology, as mentioned before. A delay component in ET path which can be read and written by Matlab instruction from control PC is implemented in FPGA. In this manner, one is able to easily sweep the temporal delay in the ET path for evaluation of the proposed methods in time domain and frequency domain.
First of all, $\tau_{\text{COM}}$ is shifted around the optimal value in order to evaluate the derived relationship between the delay mismatch and the PA output spectrum. The coefficients $k$ and $a_3$ are proven to be negative and positive for the hardware test setup, respectively. It is observed that the ACP asymmetry changes with the sweep of the delay value in ET path, as shown in Fig. 4.66. The experimental results very well match with the derived relationship between the temporal mismatch and spectrum asymmetry. With the knowledge of $a_3$, $k$ and ACP asymmetry the sign of temporal mismatch between the RF signal and its ET signal can be determined by using the Table 4.

![Figure 4.66: Measured ACP asymmetry vs. temporal mismatch.](image)

In further experiments the proposed compensation methods for the temporal mismatch are verified together with DPD and the software routine is given in Fig. 4.67. The symmetrical training signal is applied to make the first calibration before single-carrier UMTS is transmitted. Subsequently, the sign of the coefficients $k$ and $a_3$ is then estimated for the use of the Table 4. Afterwards the DPD is switched on for compensation of PA nonlinearities. Until the DPD achieves its performance limitation, namely no further improvement in terms of suppression of ACP is in sight, the proposed temporal mismatch analysis method in frequency domain is turned on to compensate the rest of the temporal misalignment between the ET signal and RF signal.
The black curve in Fig. 4.68 presents the ET system output after temporal misalignment calibration but without DPD. The green curve depicts the spectrum after running of DPD, but one can still observe the strong asymmetry of ACP. Then the time compensation using frequency domain analysis is switched on to compensate the rest of the temporal mismatch between the RF signal and its ET signal. After further adaptations of DPD additional suppression of ACP is observed on the PA output spectrum, as shown in the zoomed view in Fig 4.68.
Chapter 5 Complexity Assessment

In last decades a lot of research has been done in the field of DPD. Various DPD models such as Wiener model and Volterra model as well as different mathematical basis functions such as polynomial and orthogonal polynomial are investigated to improve the DPD performance with respect to taking into account the memory effects and strong nonlinearities. This work focuses on the computation algorithms for updating of DPD coefficients regarding the DPD performance, stability and computational cost. The developed novel methods do not directly improve the DPD performance in terms of EVM reduction and ACP suppression, but improve the stability and reduce the computational cost of DPD. Thus, more accurate and complicated DPD structures with higher order mathematical functions can be used and the DPD performance (linearization gains) can therefore be improved in this manner indirectly. In this chapter the computational cost for stationary iterative methods and the method of using adaptive selection of basis functions is discussed in comparison with conventional DPD algorithms.

5.1 Computational Cost of Stationary Iterative Methods

At first glance the computational cost of iterative methods is higher than the conventional LS method. However, through some advancement the computational cost of stationary iterative methods can be reduced significantly without any performance loss. For comparison purpose, the computational costs of conventional direct method and Gauss-Seidel method are apposed to each other. The first step of both methods is to transform the over-determined system of equations to a square system of equations, so that the direct method has same computational cost as the Gauss-Seidel method. The Gauss-Jordan method (direct method) needs $n^3/2 + n^2 - 5n/2 + 2$ complex multiplications and $n^3/2 - 3n/2 + 1$ complex additions to solve the equation (4.64), see [76], where $n$ denotes the number of DPD coefficients. For the Gauss-Seidel iterative method $(n^2 + n) \cdot l$ complex multiplications and
$(n^2 + n) \cdot l$ complex additions are required in the coefficients computation, where $l$ presents the iteration number of the iterative method.

For instance, a DPD system contains 50 coefficients. If less than 25 iterations are used in Gauss-Seidel method, the conventional direct method has even higher computational cost in comparison to Gauss-Seidel method. It has been proven in simulations and lab experiments that the iteration number for the stationary iterative methods strongly depends on the initial guess, condition number of matrix $M_y$ and the break conditions. With the tricks mentioned in Section 4.4 the number of iterations can be significantly reduced without downgrading DPD performance. Therefore, stable and good DPD performance can be achieved without increasing computational costs.

### 5.2 Computational Cost of DPD Using Adaptive Selection of Basis Functions

The second method proposed in this work for improvement of stability and reduction of computational cost is the DPD using the adaptive selection of basis functions. The direct learning DPD using adaptive selection of basis functions can significantly reduce the computational cost in conjunction with some advancement mentioned in Section 4.5.

For demonstration purpose the memory polynomial DPD with 4 functions, each functions composed of 10 basis functions, is placed under test. The polynomial with 10th order or even higher order is often applied for the PAs with strong nonlinearities, i.e. Doherty PAs. The construction of a matrix of basis functions based on the feedback signal costs $45 \cdot N$ complex multiplications. The number of feedback samples used to build the matrix $M_y$ is denoted by $N$ and normally ranges from a few thousands to tens of thousands. To perform the matrix transformation in the equation (4.64) another $N \cdot \left(n(n+1)/2 + 1\right)$ complex multiplications and $N \cdot \left((n-1)(n-2)/2 + 1\right)$ complex additions are needed, where $n$ represents the number of coefficients contained in the DPD and in this case $n$ equals 40, for 4 functions each of 10th order polynomial. The rest of the computational cost already has been given in last Section 5.1, so that the resultant computational cost of the Gauss-Jordan
method contains approximately \( N \cdot (45 + n(n + 1)/2 + 1) + n^3/2 + n^2 - 5n/2 \) complex multiplications as well as \( N \cdot ((n - 1) \cdot (n - 2)/2 + 1) + n^3/2 - 3n/2 \) complex additions.

For the DPD using adaptive selection of basis functions one can use the original signal to build the matrix \( M_{bf(n)} \). In this way \( 45 \cdot N \) complex multiplications can be saved, because the basis functions with original input signal have already been calculated in the PD and can be used directly in the DPD coefficients estimation process. However, one has to spend \( N \cdot \hat{n} \) complex multiplications and \( N \cdot \hat{n} \) complex additions to calculate the projection of the DPD residual on each basis functions.

If the residual is small enough, the DPD space \( S \) could be empty and does not contain any basis functions. Therefore, no more computational cost than \( N \cdot \hat{n} \) complex multiplications and \( N \cdot \hat{n} \) complex additions are needed. However, in most cases or without optimized threshold values, the DPD space \( S \) always contains a few basis functions. The resulting entire computational cost of this proposed method by using dynamical selection of basis functions can be given by \( N \cdot (n + \hat{n}(\hat{n} + 1)/2) + \hat{n}^3/2 + \hat{n}^2 - 5\hat{n}/2 \) complex multiplications as well as \( N \cdot (n + (\hat{n} - 1) \cdot (\hat{n} - 2)/2) + \hat{n}^3/2 - 3\hat{n}/2 \) complex additions, where \( \hat{n} \) denotes the number of selected basis functions in the current DPD space \( S \).

Assuming that 8000 samples are used in estimation of DPD coefficients, the conventional method with all 40 basis functions needs over 7 million complex multiplications and 6 million complex additions approximately. But for the method of adaptive selection of basis functions with e.g. 5 basis functions in the DPD space \( S \) one is able to cover all computation by about 448 kilo complex multiplications and 376 kilo complex additions per each coefficients adaptation. The computational cost of adaptive selection method is almost 15 times less than the computational cost of conventional DPD with all basis functions. Naturally the computational cost decreases with less basis functions in the DPD space \( S \) and thus enables faster update rates of DPD coefficients.
Chapter 6 Conclusion

This work focuses on the object of DPD and its closely related objects, such as PA modelling and signal conditioning for ET system. Several dedicated analysis theories and compensation methods are developed to address these topics. The proposed novel methods enable the DPD to deal with much stronger nonlinearities and more complicated memory effects with higher stability and less computational cost.

6.1 Contributions

Primary contributions of this work are summarized as follows:

• Designed novel method to extract the delay constant of the dominating PA memory effect by using conventional Two-Tone test.

• Applying stationary iterative methods for solving the DPD coefficients, which provides much more stable DPD performance.

• Adaptively selecting the basis functions according to the residual of DPD system and/or the requirements on PA system significantly improves the system stability and reduces the computational cost.

• Investigation of the influence of fast load change on the offline DPD system and developing the method of attaching old samples to reduce its spurious emission into neighbour bands.

• Designed novel analysis methods both in time and frequency domain as well as the compensation technique for temporal mismatch in an ET system.

• Integrated a wideband offline DPD test platform in existing hardware, which is capable of testing different DPD algorithms.
6.2 Key Results

The dominating memory delay constant can be derived from the asymmetrical IMDs by using the conventional Two-Tone test. The relationship between the PA memory effects in time domain and the PA output in frequency domain is exposed in this work.

The existing offline DPD systems with high degree of freedom are suffering from the problems of instability and high computational cost. One objective of this work is to address these two problems, i.e. to improve the DPD stability and reduce the computational cost simultaneously. The proposed novel methods of solving DPD coefficients by using stationary iterative methods and adaptive selection of basis functions achieve significant improvements in these two important aspects of DPD. By the method of using adaptive selection of basis functions the condition number of the DPD matrix, the index of reliability of DPD coefficients, can be reduced by factor of 10 or even more. Furthermore, the computational cost would be typically reduced by factor 15 at the same time.

Another weakness of offline DPD is the incapability of dealing with fast load changes because of its low update rate. The proposed method of attaching old samples to the adaptation of the new coefficients can strongly limit the spectrum regrowth, when the PA switches from low load to high or full load. This method ensures that the missing information of PA behaviour is refilled by the old samples. If PA temperature does not drift a lot during the load change, this method is capable of delivering nearly the same performance as if the PA behaviour information of high input power level would not be missing. If PA temperature slightly follows the load change, this method can still provide satisfying performance. Furthermore, this method does not have any additional signal process in the feedforward path, so that no extra latency time is introduced with this method. Moreover, this method needs very small modification of existing DPD adaptation process, which can be easily integrated in existing DPD systems.

The temporal misalignment between PA RF input signal and its ET signal has a large impact on the signal conditioning processes in the ET system. The compensation methods with and
without training signal are developed, thereby the temporal misalignment can be calibrated at the first with a special training signal and continuously adjusted during the operating time.

6.3 Suggestions for Future Research

The future research can be applied in the following directions:

- Extension of adaptive selection of basis functions to dynamic adaptation of basis functions.
- Increasing the linearization bandwidth with the extension of more memory functions and cross terms.
- Extension of DPD structures and algorithms to enable multi-band transmission by using one PA.
Appendix A. Memory Delay Estimation for PA with Memory Polynomial Model

The deviation of the delay constant for dominating PA memory effect is based on memory polynomial RF model with only one memory functions, as given by:

\[ y(t) = |a_{1,0}| \cdot \text{real}\left(e^{j\varphi_{1,0}} \cdot \tilde{x}(t)\right) + |a_{3,0}| \cdot \text{real}\left(e^{j\varphi_{3,0}} \cdot \tilde{x}(t)\right)^3 + |a_{5,0}| \cdot \text{real}\left(e^{j\varphi_{5,0}} \cdot \tilde{x}(t)\right)^5 + \cdots \]

\[ + |a_{1,1}| \cdot \text{real}\left(e^{j\varphi_{1,1}} \cdot \tilde{x}(t-\tau)\right) + |a_{3,1}| \cdot \text{real}\left(e^{j\varphi_{3,1}} \cdot \tilde{x}(t-\tau)\right)^3 + |a_{5,1}| \cdot \text{real}\left(e^{j\varphi_{5,1}} \cdot \tilde{x}(t-\tau)\right)^5 + \cdots \]  \( (a.1) \)

where the |\( \cdot |\) denotes the absolute value operation. \( a_{n,m} \) and \( \varphi_{n,m} \) describe the complex coefficients and their angles, respectively. The Two-Tone test signal at angle frequency \( \omega_1 \) and \( \omega_2 \) can be expressed as:

\[ x(t) = \text{real}\left(A \cdot e^{j\omega_1 t} + A \cdot e^{j\omega_2 t}\right) \]  \( (a.2) \)

\[ \tilde{x}(t) = A \cdot e^{j\omega_1 t} + A \cdot e^{j\omega_2 t} \]

\[ x(t-\tau) = \text{real}\left(A \cdot e^{j(\omega_1 t-\omega_2 \tau)} + A \cdot e^{j(\omega_2 t-\omega_1 \tau)}\right) \]  \( (a.3) \)

\[ \tilde{x}(t-\tau) = A \cdot e^{j(\omega_1 t-\omega_2 \tau)} + A \cdot e^{j(\omega_2 t-\omega_1 \tau)} \]

where \( x(t) \) represents the real part of \( \tilde{x}(t) \). The complex signal \( \tilde{x}(t) \) is only the auxiliary signal for simplification of mathematical derivation. Under the assumption that only the \( n^{th} \) order polynomial term has contribution to the \( n^{th} \) order IMD and hence the \( n^{th} \) distortion can be defined by:

\[ D_n = |a_{n,0}| \cdot \text{real}\left(e^{j\varphi_{n,0}} \cdot \tilde{x}(t)\right)^n + |a_{n,1}| \cdot \text{real}\left(e^{j\varphi_{n,1}} \cdot \tilde{x}(t-\tau)\right)^n. \]  \( (a.4) \)
Because only the power of the test tones is taken account in Two-Tone test, one can rotate the \( n \)th order distortion about \( n \) to simplify the rest derivation, which leads to

\[
D_n^{\text{rotation}} = |a_{n,0}| \cdot \text{real}(\tilde{x}(t))^n + |a_{n,1}| \cdot \text{real} \left( e^{j \frac{\varphi_{n,0} - \varphi_{n,1}}{n}} \cdot \tilde{x}(t - \tau) \right)^n. \tag{a.5}
\]

At first only the IMD3 distortion is in focus:

\[
\text{IMD3}^{\text{rotation}} = \frac{3}{4} |a_{3,0}| A^3 \left[ \cos(2\omega t - \omega_3 t) + \cos(2\omega t - \omega t) \right] \\
+ \frac{3}{4} |a_{3,1}| A^3 \left[ \cos \left( 2\omega t - \omega_3 t + 2\omega t - \omega_3 t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right] \\
+ \cos \left( 2\omega t - \omega t + 2\omega t - \omega t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right). \tag{a.6}
\]

The IMD3 distortion can be further simplified by using trigonometric addition and subtraction formula, as given by:

\[
\text{IMD3}^{\text{rotation}} = \left[ \frac{3}{4} |a_{3,0}| A^3 + \frac{3}{4} |a_{3,1}| A^3 \cos \left( 2\omega t - \omega_3 t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right] \cdot \cos(2\omega t - \omega_3 t) \\
- \frac{3}{4} |a_{3,1}| A^3 \cdot \sin \left( 2\omega t - \omega t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \cdot \sin(2\omega t - \omega_3 t) \\
+ \left[ \frac{3}{4} |a_{3,0}| A^3 + \frac{3}{4} |a_{3,1}| A^3 \cos \left( 2\omega t - \omega t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right] \cdot \cos(2\omega t - \omega_3 t) \\
- \frac{3}{4} |a_{3,1}| A^3 \cdot \sin \left( 2\omega t - \omega t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \cdot \sin(2\omega t - \omega_3 t). \tag{a.7}
\]

Subsequently, the power of lower and upper IMD3 can be presented by:

\[
\text{IMD3}_{\text{Low-Pwr}} = \left[ \frac{3}{4} |a_{3,0}| A^3 + \frac{3}{4} |a_{3,1}| A^3 \cos \left( 2\omega t - \omega_3 t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right]^2 \\
+ \left[ \frac{3}{4} |a_{3,1}| A^3 \cdot \sin \left( 2\omega t - \omega_3 t + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right]^2. \tag{a.8}
\]
\[ \text{IMD3\_Up\_Pwr} = \left[ \frac{3}{4} |a_{3,0}| \cdot A^2 + \frac{3}{4} |a_{3,1}| \cdot A^3 \cos \left( 2\omega_2 \tau - \omega_1 \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right]^2 + \left[ \frac{3}{4} |a_{3,1}| \cdot A^3 \cdot \sin \left( 2\omega_2 \tau - \omega_1 \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right]^2, \] (a.9)

with the assumption \( \omega_1 < \omega_2 \).

At the end the resultant formula for the two IMD3s can be given by:

\[ \text{IMD3\_Low\_Pwr} = \frac{9}{16} |a_{3,0}| \cdot A^6 + \frac{9}{8} |a_{3,1}| \cdot A^6 \]
\[ + \frac{9}{8} |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cos \left( \omega \tau - 3\Delta \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \] (a.10)

\[ \text{IMD3\_Up\_Pwr} = \frac{9}{16} |a_{3,0}| \cdot A^6 + \frac{9}{8} |a_{3,1}| \cdot A^6 \]
\[ + \frac{9}{8} |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cos \left( \omega \tau + 3\Delta \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \] (a.11)

with \( \omega = \frac{\omega_1 + \omega_2}{2}; \Delta = \left| \frac{\omega_1 - \omega_2}{2} \right| \),

where \( \omega \) and \( \Delta \) denote the central angle frequency and half angle frequency spacing of the two-tone signal, respectively. One can calculate the difference of the IMD3s by:

\[ \text{IMD3\_Low\_Pwr} - \text{IMD3\_Up\_Pwr} = \frac{9}{8} |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \left( \cos \left( \omega \tau - 3\Delta \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) - \cos \left( \omega \tau + 3\Delta \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \right) \]
\[ = \frac{9}{4} |a_{3,0}| \cdot |a_{3,1}| \cdot A^6 \cdot \sin \left( \omega \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \cdot \sin(3\Delta \tau) \] (a.12)

In equation (a.12) the term \( \sin(3\Delta \tau) \) is dependent on the frequency spacing of the two-tone signal \( \Delta \) and the memory delay constant \( \tau \). Therefore, if one sweeps the two-tone signal with a constant central frequency \( \omega \), the last second term \( \sin \left( \omega \tau + \frac{\varphi_{3,1} - \varphi_{3,0}}{3} \right) \) would
remain unchanged. Theoretically one needs only two result sets to solve the time delay constant \( \tau \), as given by:

\[
\begin{align*}
\frac{\text{IMD}_{\text{diff}} (\Delta_1)}{\text{IMD}_{\text{diff}} (\Delta_2)} &= \frac{\frac{9}{4} \left| a_{3,1} \right| \cdot A^6 \cdot \sin \left( \frac{\omega \tau + \frac{\phi_{3,1} - \phi_{3,0}}{3}}{3} \right) \sin (3\Delta_1 \tau)}{\frac{9}{4} \left| a_{3,0} \right| \cdot A^6 \cdot \sin \left( \frac{\omega \tau + \frac{\phi_{3,1} - \phi_{3,0}}{3}}{3} \right) \sin (3\Delta_2 \tau)} \\
&= \frac{\sin (3\Delta_1 \tau)}{\sin (3\Delta_2 \tau)} = G. 
\end{align*}
\]  

(a.13)

To further simplify the equation one could double the frequency spacing of first test by the second test, which leads the equation (a.13) to:

\[
\begin{align*}
\frac{\text{IMD}_{\text{diff}} (\Delta_1)}{\text{IMD}_{\text{diff}} (\Delta_2)} &= \frac{\sin (3\Delta_1 \tau)}{\sin (3\Delta_2 \tau)} = \frac{\sin (3\Delta_1 \tau)}{\sin (6\Delta_1 \tau)} = \frac{\sin (3\Delta_1 \tau)}{2 \sin (3\Delta_1 \tau) \cdot \cos (3\Delta_1 \tau)} = G \\
\Rightarrow \tau &= \frac{2k \pi \pm \arccos \left( \frac{1}{2G} \right)}{3 \Delta_1}; k = \{0, \pm 1, \pm 2 \cdots \}. 
\end{align*}
\]  

(a.14)

with \( \Delta_2 = 2\Delta_1 \).

Similarly, one is able to derive the memory delay constant from the asymmetry of all these IMDs as well as the two fundamental tones. Under the same conditions of constant central frequency of successive tests and doubled frequency spacing of the second try, one can find a simple and aesthetic mathematical formula,

\[
\tau = \frac{\arccos \left( \frac{1}{2G} \right)}{n \cdot \Delta_1}
\]  

(a.15)

\( n = 1, 3, 5, 7 \ldots \) (odd number)

to describe the relationship between the delay constant \( \tau \) of the dominating PA memory effect and the asymmetry of the IMDs. The variable \( n \) denotes the order of the IMDs, e.g., 1 for fundamental tone, 3 for third order IMD, 5 for fifth order IMD and so on.
Appendix B. Proof of Positive-Semi-Definite Matrix

Definition 1: Matrix $M$ is positive-semi-definite, if and only if $z^H \cdot M \cdot z \geq 0$ with all $z \in \mathbb{C}^n$ and $z \neq 0$.

The matrix $(M_y^H \cdot M_y)$ is a positive-semi-definite matrix.

Proof: $z \in \mathbb{C}^n$ and $z^H \cdot (M^H \cdot M) \cdot z \geq 0$

$z^H \cdot (M^H \cdot M) \cdot z$

$= (z^H \cdot M^H_y) \cdot (M_y \cdot z)$

$= (M_y \cdot z)^H \cdot (M_y \cdot z)$ with $(z^H \cdot M^H_y) = (M_y \cdot z)^H$

$= \sum_{i=1}^{n} |a_i|^2 \geq 0$; with $(M_y \cdot z) = [a_1, a_2, \ldots a_n]^T$, $a_i \in \mathbb{C}$
Appendix C. Relationship between Temporal Mismatch and Unsymmetrical IMDs

The mathematical description for a memoryless PA base band model with only third order distortion is given by:

\[ y[n] = x[n] \cdot \left( a_1 + a_3 \cdot |x[n]|^2 \right) e^{j \phi(|x[n]|)} \]  

(c.1)

with \( a \in R \),

whereby the phase distortion is defined by [75]:

\[ \phi(|x[n]|) = 2k \cdot |x[n]|^2. \]  

(c.2)

Two-tone signal \( \cos(\omega n) \) is used as stimulus for this PA model and resultant PA output signal \( y[n] \) is given by:

\[ y[n] = \left( a_1 + a_3 \cdot \cos^2(\omega n + \tau_{AM}) \right) \cdot e^{j (2k \cos^2(\omega n))} \cdot \cos(\omega n), \]  

(c.3)

where \( \tau_{AM} \) is introduced to model the delay effect caused by the temporal mismatch of ET signal. By using the double-angle formula

\[ \cos(2\theta) = 2\cos^2(\theta) - 1, \]  

(c.4)

the equation (c.3) can be written as:

\[ y[n] = \left( a_1 + a_3 \cdot \cos^2(\omega n + \tau_{AM}) \right) \cdot e^{j (k + k \cos^2(\omega n))} \cdot \cos(\omega n) \]

\[ = \left( \left( a_1 + \frac{a_3}{2} \right) + \frac{a_3}{2} \cdot 2 \cos(2\omega n + 2\tau_{AM}) \right) \cdot e^{j (k + k \cos^2(\omega n))} \cdot \cos(\omega n). \]  

(c.5)

Since only the amplitude dependent phase distortion is of interest, rotation of PA output by \( e^{-j k} \) leads to
\[ y[n] = \left( a_1 + \frac{a_3}{2} \right) + \frac{a_3}{2} \cdot \cos(2\omega n + 2\tau_{AM}) \cdot e^{j k \cos 2\omega n} \cdot \cos(\omega n). \] (c.6)

If the value of \( k \) is very small, the AM-PM distortion described in equation (c.6) can be approximately given by:

\[
e^{j k \cos 2\omega n} = \cos(k \cdot \cos(2\omega n)) + j \cdot \sin(k \cdot \cos(2\omega n)) = 1 - k \cdot \cos(2\omega n) + j \cdot k \cdot \cos(2\omega n) \] (c.7)

with \( k \to 0 \).

Therefore, the PA output can be expressed as:

\[
y'[n] = \left( a_1 + \frac{a_3}{2} \right) + \frac{a_3}{2} \cdot \cos(2\omega n + 2\tau_{AM}) \cdot (1 - k \cdot \cos(2\omega n) + j \cdot k \cdot \cos(2\omega n)) \cdot \cos(\omega n). \] (c.8)

The IMD3 of interest is given by:

\[
y''_3 = \left( a_1 + \frac{a_3}{2} \right) \cdot \cos(3\omega n + 2\tau_{AM}) - \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} \cdot \cos(3\omega n)
+ j \cdot \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} \cdot \cos(3\omega n) + \frac{a_3 k}{8} \cdot \cos(3\omega n + 2\tau_{AM}) \right). \] (c.9)

The equation (c.9) can be further simplified by using trigonometric addition and subtraction formula:

\[
y''_3[n] = \left( \frac{a_1}{4} \cdot \cos(2\tau_{AM}) - \frac{a_3 k}{8} \cdot \cos(2\tau_{AM}) - \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} \cdot \frac{e^{j 3\omega n} + e^{-j 3\omega n}}{2} 
- \left( \frac{a_3}{4} \cdot \sin(2\tau_{AM}) - \frac{a_3 k}{8} \cdot \sin(2\tau_{AM}) \right) \cdot \frac{e^{j 3\omega n} - e^{-j 3\omega n}}{2j} 
+ j \left( \left( a_1 + \frac{a_3}{2} \right) \cdot \frac{k}{2} + \frac{a_3 k}{8} \cdot \cos(2\tau_{AM}) \right) \cdot \frac{e^{j 3\omega n} + e^{-j 3\omega n}}{2} - \frac{a_3 k}{8} \cdot \sin(2\tau_{AM}) \cdot \frac{e^{j 3\omega n} - e^{-j 3\omega n}}{2j} \right) \] (c.10)

with \( \cos(3\omega n) = \frac{e^{j 3\omega n} + e^{-j 3\omega n}}{2} \) and \( \sin(3\omega n) = \frac{e^{j 3\omega n} - e^{-j 3\omega n}}{2j} \).

The reform of the equation (c.10) in upper IMD3 and lower IMD3 leads to
Because only the power of IMD3 is of interest and the real parts for the both IMD3 are identical, one merely needs to compare the imaginary parts to find out which power is larger than the other. The imaginary parts of upper and lower IMD3 are given by:

\[
IMD_{3\_U_{im}} = \left( a_1 + \frac{a_3}{2} + \frac{a_4}{4} \cdot \cos(2\tau_{AM}) \right) \cdot \frac{k}{2} + \frac{a_3}{4} \sin(2\tau_{AM})
\]

and

\[
IMD_{3\_L_{im}} = \left( a_1 + \frac{a_3}{2} + \frac{a_4}{4} \cdot \cos(2\tau_{AM}) \right) \cdot \frac{k}{2} - \frac{a_3}{4} \sin(2\tau_{AM})
\]

respectively.
List of References


