THE UNIVERSITY OF ELECTRO-COMMUNICATIONS

# Advanced Digital Signal Processing Technology for Nonlinear Compensation in Wireless Communication Systems

---

Department of Communication Engineering and Informatics

**Yuelin Ma**
**2015/2/25**

*A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy*
March 2015

# Advanced Digital Signal Processing Technology for Nonlinear Compensation in Wireless Communication Systems

SUPERVISORY COMMITTEE:

Professor Yasushi Yamao

Professor Kazuhiko Honjo

Professor Yoshio Karasawa

Professor Takeo Fujii

Associate Professor Toshiharu Kojima

# 概要

一般に電子デバイスを用いた無線回路は非線形特性を有するため，信号にひずみが生じる．特に送信機および受信機のフロントエンドの非線形はシステムの性能に大きな影響を与えるため，非線形補償技術が極めて重要である．本研究は，高度なディジタル信号処理を含む非線形補償のための適応信号処理技術の理論的検討とその実現に関する研究である．理論解析および汎用測定器による実験に加えて，高速DA／AD変換器も含めた実際の回路(FPGA，DSPなど)での広帯域信号処理実装を研究対象に含む．

本論文の主な貢献は以下のとおりである：

１）マイクロ波電力増幅器の非線形補償は携帯電話基地局装置の電力効率向上の面から重要な課題である．特にピーク電力の大きなマルチキャリア信号の増幅では，電力効率の面からドハティ増幅器が有効であるが，複雑な非線形特性を有するドハティ増幅器に対して安定かつ高精度な非線形補償を可能とすることは困難であり，特に隣接チャネル漏洩電力規格の厳しいGSMシステム送信機用へのディジタルプリディストーション（Digital Predistortion; DPD）方式の適用はこれまで報告されていない．本文ではこの非線形補償を実現するため，新しいアルゴリズムを考案し，0.2ms以下の収束速度を実現した．またメモリ効果を有する複雑な非線形を補償するために複数の非線形補償テーブルを用いつつ，アルゴリズムの最適化によって回路規模を抑え，FPGAへの簡易な実装を可能とし，基地局装置のコスト低減を可能した．

２）DPDは送信回路と受信回路が含まれ，受信回路は一般に送信回路より重要である．受信回路はひずみによって帯域幅の広くなった非線形信号を受信するため，帯域幅が十分大きい必要がある．そのため，高速ADCと高速信号処理が必要になり，コストが非常に高くなるという問題がある．今後の導入が期待される第５世代（5G）無線通信で必要となる，超広帯域信号用DPDの性能向上とコストの低減を同時に満たすため，非線形信号の帯域幅に比べて狭帯域のAD変換器を用いて信号を処理する技術を提案した．この手法により，標本化周波数が300 MHzのAD変換器を用いながら，320MHz以上の信号帯域幅を有する信号の非線形補償を実現した．

３）無線通信のトラフィックの増大は著しく，第４世代の移動通信システムLTE-Advancedでは複数の帯域を同時に使用するCarrier Aggregationの導入が決定されている．この場合，送信機と受信機は異なる帯域にある複数キャリアを同時に送信・受信することになる．第４世代以降で本格導入される同時複数帯域

（concurrent multi-band）送受信技術では，フロントエンドが複数帯域で共有されるので，回路の非線形性による相互変調ひずみの影響が複雑かつ顕著となる．本研究では，送信機と受信機に対して，それぞれ効果的にひずみを低減できる複数帯域同時非線形補償技術を提案し，その効果を実験により実証している．

**キーワード:** 非線形ひずみ，線形化，非線形補償，ディジタルプリディストータ，ADC，FPGA，LTE，LTE-Advanced，GSM，フロントエンド，concurrent dual-band，相互帯変調

# Summary

The inherit nonlinearity in analogue front-ends of transmitters and receivers have had primary impact on the overall performance of the wireless communication systems, as it gives arise of substantial distortion when transmitting and processing signals with such circuits. Therefore, the nonlinear compensation (linearization) techniques become essential to suppress the distortion to an acceptable extent in order to ensure sufficient low bit error rate. Furthermore, the increasing demands on higher data rate and ubiquitous interoperability between various multi-coverage protocols are two of the most important features of the contemporary communication system. The former demand pushes the communication system to use wider bandwidth and the latter one brings up severe coexistence problems.

Having fully considered the problems raised above, the work in this Ph.D. thesis carries out extensive researches on the nonlinear compensations utilizing advanced digital signal processing techniques. The motivation behind this is to push more processing tasks to the digital domain, as it can potentially cut down the bill of materials (BOM) costs paid for the off-chip devices and reduce practical implementation difficulties. The work here is carried out using three approaches: numerical analysis & computer simulations; experimental tests using commercial instruments; actual implementation with FPGA.

The primary contributions for this thesis are summarized as the following three points:

1) An adaptive digital predistortion (DPD) with fast convergence rate and low complexity for multi-carrier GSM system is presented. Albeit a legacy system, the GSM, however, has a very strict requirement on the out-of-band emission, thus it represents a much more difficult hurdle for DPD application. It is successfully implemented in an FPGA without using any other auxiliary processor. A simplified multiplier-free NLMS algorithm, especially suitable for FPGA implementation, for fast adapting the LUT is proposed. Many design methodologies and practical implementation issues are discussed in details. Experimental results have shown that the DPD performed robustly when it is involved in the multichannel transmitter.

2) The next generation system (5G) will unquestionably use wider bandwidth to support higher throughput, which poses stringent needs for using high-speed data converters. Herein the analog-to-digital converter (ADC) tends to be the most expensive single device in the whole transmitter/receiver systems. Therefore, conventional DPD utilizing high-speed ADC becomes unaffordable, especially for small base stations (micro, pico and femto). A digital predistortion technique utilizing spectral extrapolation is proposed in this thesis, wherein with band-limited feedback signal, the requirement on ADC speed can be significantly released. Experimental results have validated the feasibility of the proposed technique for coping with band-limited feedback signal. It has been shown that adequate linearization performance can be achieved even if the acquisition bandwidth is less than the original signal bandwidth. The experimental results obtained by using LTE-Advanced signal of 320 MHz

bandwidth are quite satisfactory, and to the authors' knowledge, this is the first high-performance wideband DPD ever been reported.

3) To address the predicament that mobile operators do not have enough contiguous usable bandwidth, carrier aggregation (CA) technique is developed and imported into 4G LTE-Advanced. This pushes the utilization of concurrent dual-band transmitter/receiver, which reduces the hardware expense by using a single front-end. Compensation techniques for the respective concurrent dual-band transmitter and receiver front-ends are proposed to combat the inter-band modulation distortion, and simultaneously reduce the distortion for the both lower-side band and upper-side band signals.

**Keywords:** nonlinear distortion, linearization, nonlinear compensation, digital predistortion, ADC, FPGA, LTE, LTE-Advanced, GSM, front-end, concurrent dual-band, cross-band modulation, 5G, wideband, broadband

# Acknowledgements

Reminiscing my doctor's work here in AWCC and my whole 20 years' student life, all sorts of feelings well up in my mind. It was really a hard choice to pursue a Ph.D. degree, as it means to give up the chance to earn some real money in a steady job and, in a sense, lose the chance to gain social experience. However, the wish to achieve something significant and to discover something new impelled me finally make the decision. The results turn out to be good, and I realize that I have developed tremendously, not just as a fulfilled student, but also as a better person. Therefore, I wish to express my sincere gratitude to those who provided assistance and support through this work.

First, I would like to thank my research advisor, Professor Yasushi Yamao, for guiding me throughout my work. His brilliance, unpretentious, enthusiasm and attitude support not only brings out the best in his students, but also provides us with a good model to follow. I was occasionally astonished by his high comprehensive ability, as I couldn't describe things correctly due to my incompetent Japanese. Without his guidance, my work wouldn't have been so fruitful and I am truly indebted to him.

I would thank Professor Yoshihiko Akaiwa for enlightening me and leading me into the way of doing research. It was his enlightenment have been my great interests in research stimulated. I will never forget his one-to-one teaching in his office when I was an exchange student in AWCC.

I would like to thank JSPS (Japan Society for the Promotion of Science) research fellowship for young scientists funding that provides me financial support, so that I can survive in the world most expensive city.

Finally, I want to share my excitement with my family, and thank their love and support.

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| 2G | **2**nd **G**eneration |
| 3G | **3**rd **G**eneration |
| 4G | **4**th **G**eneration |
| 5G | **5**th **G**eneration |
| | |
| ACLR | **A**djacent **C**hannel **L**eakage **R**atio |
| ACP | **A**djacent **C**hannel **P**ower |
| AGC | **A**utomatic **G**ain **C**ontrol |
| AM-AM | **Am**plitude-to-**A**mplitude **M**odulation |
| AM-PM | **Am**plitude-to-**P**hase **M**odulation |
| AWG | **A**rbitrary **W**ave **G**enerator |
| | |
| BOM | **B**ill **o**f **M**aterials |
| BPF | **B**and**p**ass **F**ilter |
| BS | **B**ase **S**tation |
| | |
| CA | **C**arrier **A**ggregation |
| CCDF | **C**omplementary **C**umulative **D**istribution **F**unction |
| CFR | **C**rest **F**actor **R**eduction |
| CP | **C**yclic **P**refix |
| CW | **C**ontinuous **W**ave |
| | |
| DAC | **D**igital-to-**A**nalog **C**onverter |
| DDC | **D**igital **D**own **C**onverter |
| DDS | **D**irect **D**igital **S**ynthesizer |
| DFE | **D**igital **F**ront-**E**nd |
| DPD | **D**igital **P**re**d**istortion |
| DSP | **D**igital **S**ignal **P**rocessor |
| DUC | **D**igital **U**p **C**onverter |
| DUT | **D**evice **U**nder **T**est |
| | |
| EDGE | **E**nhanced **D**ata Rates for **GSM** **E**volution |
| EVM | **E**rror **V**ector **M**agnitude |

| | |
|---|---|
| | |
| FDD | **F**requency **D**ivision **D**uplex |
| FET | **F**ield **E**ffect **T**ransistor |
| FFT | **F**ast **F**ourier **T**ransform |
| FIR | **F**inite **I**mpulse **R**esponse |
| FMC | **F**PGA **M**ezzanine **C**ard |
| FoM | **F**igure **o**f **M**erit |
| FPGA | **F**ield **P**rogrammable **G**ate **A**rray |
| | |
| GPIB | **G**eneral **P**urpose **I**nterface **B**us |
| GPRS | **G**eneral **P**acket **R**adio **S**ervice |
| GSM | **G**lobal **S**ystem for **M**obile Communications |
| | |
| HPC | **H**igh **P**in **C**onnector |
| HPF | **H**igh **P**ass **F**ilter |
| | |
| IBO | **I**nput **B**ack**o**ff |
| IF | **I**ntermediate **F**requency |
| IFFT | **I**nverse **F**ast **F**ourier **T**ransform |
| IM | **I**nter**m**odulation |
| IMD | **I**nter**m**odulation **D**istortion |
| ISI | **I**nter **S**ymbol **I**nterference |
| | |
| LAN | **L**ocal **A**rea **N**etwork |
| LINC | **Li**near Amplification using **N**onlinear **C**omponents |
| LMS | **L**east **M**ean **S**quares |
| LNA | **L**ow **N**oise **A**mplifier |
| LO | **L**ocal **O**scillator |
| LPF | **L**ow **P**ass **F**ilter |
| LS | **L**east **S**quares |
| LSB | **L**ower **S**ide **B**and |
| LTE | **L**ong **T**erm **E**volution |
| LTE-A | **L**ong **T**erm **E**volution **A**dvanced |
| | |
| MC-GSM | **M**ulti-**C**arrier **GSM** |
| MSE | **M**ean **S**quare **E**rror |
| | |
| NLC | **N**on**l**inear **C**ompensation |

| | |
|---|---|
| NLMS | **N**ormalized **L**east **M**ean **S**quare |
| NMSE | **N**ormalized **M**ean **S**quare **E**rror |
| | |
| OBO | **O**utput **B**ack**o**ff |
| OFDM | **O**rthogonal **F**requency **D**ivision **M**ultiplexing |
| OSR | **O**ver**s**ampling **R**atio |
| | |
| PA | **P**ower **A**mplifier |
| PAE | **P**ower **A**dded **E**fficiency |
| PAPR | **P**eak to **A**verage **P**ower **R**atio |
| PC | **P**eak **C**ancellation |
| PD | **P**re**d**istorter |
| PDF | **P**robability **D**ensity **F**unction |
| PM | **P**hase **M**odulated |
| PNLC | **P**ost **N**on**l**inear **C**ompensation |
| PSD | **P**ower **S**pectral **D**ensity |
| | |
| QDMOD | **Q**uadrature **D**e**mod**ulator |
| QMOD | **Q**uadrature **Mod**ulator |
| | |
| RAM | **R**andom **A**ccess **M**emory |
| RF | **R**adio **F**requency |
| ROM | **R**ead **O**nly **M**emory |
| RRC | **R**oot **R**aised **C**osine |
| | |
| SA | **S**ignal **A**nalyzer |
| SCPI | **S**tandard **C**ommands for **P**rogrammable **I**nstruments |
| SDR | **S**purious **D**ynamic **R**ange |
| SFFB | **S**pectra**-F**olding **F**eedback |
| SG | **S**ignal **G**enerator |
| SoC | **S**ystem **o**n a **C**hip |
| SPI | **S**erial **P**eripheral **I**nterface |
| SVD | **S**ingular **V**alue **D**ecomposition |
| | |
| UI | **U**nit **I**nterval |
| USB | **U**pper **S**ide **B**and |
| | |
| VCCS | **V**oltage **C**ontrol **C**urrent **S**ource |

| | |
|---|---|
| VGA | **V**ariable **G**ain **A**mplifier |
| VISA | **V**irtual **I**nstrument **S**oftware **A**rchitecture |
| VSA | **V**ector **S**ignal **A**nalyzer |
| VSG | **V**ector **S**ignal **G**enerator |
| | |
| WCDMA | **W**ideband **C**ode **D**ivision **M**ultiple **A**ccess |
| WLAN | **W**ireless **L**ocal **A**rea **N**etwork |
| | |
| XOR | E**x**clusive **OR** |

# Glossary

| | |
|---|---|
| $\bullet^*$ | Conjugate |
| $\bullet^T$ | Matrix transpose |
| $\bullet^H$ | Hermit transpose |
| $\angle\bullet$ | Angle |
| $\tilde{\bullet}$ | Signal in passband |
| $\lvert\bullet\rvert$ | Absolute for real number |
| $\lvert\lvert\bullet\rvert\rvert$ | Magnitude of complex number |
| $\lvert\lvert\bullet\rvert\rvert_2$ | 2-norm |
| $\mathbf{a}$ | Coefficient vector containing $a_i$ |
| $A(t)$ | Time variant amplitude |
| $a_i$ | Polynomial coefficients |
| B | Normalized bandwidth |
| $\mathrm{B}_l$ | Normalized bandwidth for band-limited acquisition path |
| $\mathbf{b}$ | Coefficient vector containing $b_i$ |
| $\mathbb{C}^N$ | $N$-dimensional complex subspace |
| $b_i$ | Polynomial coefficients |
| $\mathrm{d}_f$ | Fractional delay |
| $\mathrm{d}_i$ | Integer delay |
| $\mathbf{d}$ | Desired signal in vector form |
| $D[\ ]$ | Difference function |
| $e$ | Exponential |
| $err$ | Error |
| $E\{\ \}$ | Expectation |
| $f(\ )$ | A nonlinear function |

| | |
|---|---|
| $f_p(\ )$ | Post-inverse of $f(\ )$ |
| $f^{-1}(\ )$ | Inverse of $f(\ )$ |
| $f_c$ | Center frequency |
| $\mathbf{G}$ | Coefficient vector for a desired gain |
| $g(\ )$ | A nonlinear function |
| $g_i$ | Power series coefficients |
| $G_i(\ )$ | Look-up-table function |
| $G_{lin}$ | Linear gain |
| $\mathbf{h}$ | Filter coefficient vector |
| $\mathbf{h}_c$ | Compensation filter coefficient vector |
| $\mathbf{H}$ | Filter coefficient in Toeplitz form |
| $h_p(\ )$ | $p$th-order Volterra kernel |
| $j$ | Imaginary unit, $j^2$=-1 |
| $J(\ )$ | Cost function |
| $k_i$ | Kernel function |
| $K(\ )$ | Kernel space |
| $L$ | Vector length |
| $m$ | Memory index |
| $M$ | Memory order |
| $m(t)$ | Multi-tone signal |
| $n$ | Time index |
| $N$ | Nonlinear order |
| $P$ | A given constant, possibly means a nonlinear order |
| $\mathbf{P}$ | Coefficients matrix after polynomial multiplication |
| $P(\ )$ | Instantaneous power |
| $PAPR(\ )$ | Taking the PAPR |
| $P_{\text{consume}}$ | Consumed power |
| $P_{\text{output}}$ | Output Power |

| | |
|---|---|
| $P_{\text{supply}}$ | Supply Power |
| $PM$ | Phase shift |
| $\mathbf{p}_m$ | Coefficients vector after polynomial multiplication |
| $\mathbf{P}_i$ , $\mathbf{P}_u$ | Projection operators |
| $Q$ | A given constant, possibly means a nonlinear order |
| $\mathbb{R}^N$ | $N$-dimensional real subspace |
| $\Re\{\}$ | Real part |
| $R(\ )$ | Error function |
| $\text{ran}(\ )$ | Range |
| $sign(\ )$ | Sign function |
| $s_r$ | Reference symbol |
| $s_t$ | Test Symbol |
| $\mathbf{U}$ | Rejection function that orthogonalize a space |
| $u(t)$ | An intact signal |
| $v(n)$ | Intermedia signal, normally a nonlinear function output |
| $w_i$ | Coefficients |
| $\mathbf{W}$ | Coefficient vector containing $w_i$ |
| $\mathbf{W}_N$ | $N$-point FFT transform matrix |
| $\mathbf{X}$ | Model space |
| $x(t)$ | Baseband signal |
| $\tilde{x}(t)$ | Passband signal |
| $X(\omega)$ | Frequency domain signal |
| $x_L(t)$ | Signal in lower side band |
| $x_H(t)$ | Signal in upper side band |
| $\alpha$ | Parameter for $\alpha$Max + $\beta$Min algorithm |
| $\beta$ | Parameter for $\alpha$Max + $\beta$Min algorithm |
| $\beta_{qr}$ | Parameter used for deembedding |
| $\gamma$ | Step size |

| | |
|---|---|
| $\theta(t)$ | Time variant phase |
| $\tau$ | Time delay |
| $\delta$ | Smooth factor for exponential moving averaging |
| $\omega$ | primitive $N$-th root of unity |
| $\omega_c$ | Center frequency in radium |
| $\mu$ | Parameter for $\mu$-law |
| $\eta$ | Power efficiency |
| $\varphi_i$ | Orthogonal subspace |
| $\boldsymbol{\Psi}$ | Orthogonal space |
| $\boldsymbol{\Xi}$ | A projector recovering the null space |
| $\boldsymbol{\Omega}$ | Composited polynomial space |

# Chapter 1.

# Introduction

Contemporary wireless communication systems have been experiencing rapid evolutions and are keeping evolving in order to satisfy the increasing demands on data communication services. Although digital circuits continue to scale in the wireless system, there still remains a portion that is taken up by analog devices. The most commonly used active analog devices are transistors and diodes, which however, show nonlinearity with their respective dissimilar behaviors.

In this thesis, we consider the nonlinearities existing in wireless transmitter (TX) and receiver (RX) front-ends, which mainly originate from the amplifiers (including power amplifier and low noise amplifier), as they yield considerable amount of distortion that significantly deteriorates the overall performance of the TX/RX systems. The nonlinear devices, needless to say, distort the signal and most importantly expand the signal spectrum several folds wider, see Fig. 1.1. The distortion, with much wider spectrum, falls onto the intact signal, causing both in-band and out-of-band interferences. Generally, such nonlinear distortion occurs when the input exceeds a certain marginal power of the nonlinear device, which means larger the input power the higher probability of the occurrence of nonlinearity.

## 1.1 BACKGROUND AND MOTIVATION - THE NEED FOR NONLINEAR COMPENSATION IN WIRELESS COMMUNICATION SYSTEMS

Due to limited available frequency resource and wireless regulations, a prohibitive provision is set to the transmitters for stipulating the maximum amount of out-of-band emission,

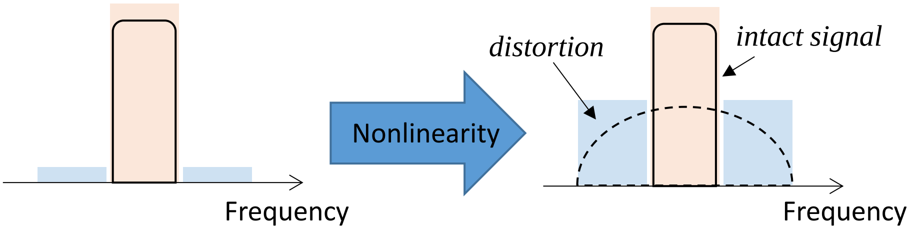


**Fig. 1.1 Demonstration of the nonlinear distorted spectrum.**

as it causes severe interference to the adjacent channels used by other receivers. Nonlinearity also has a direct impact on the receiver when an invading signal with high power appears adjacent to the channel in use. The result is the expanded spectrum of the invading signal results from nonlinear devices, which interferes the desired signal and in some circumstances even cause inter-band modulation distortion. More details about the nonlinearity appearing in transmitter and receiver, as well as the corresponding remedies are introduced in Section 1.1.1 and 1.1.2.

The prevalence of high-level modulation signal poses a strict requirement on the amount of distortion introduced by the TX and RX circuitries, forcing the use of digital compensation to facilitate such goal. Furthermore, the increasing demands on higher data rate and ubiquitous interoperability between various multi-coverage protocols lead to the development of heterogeneous networks. The former demand justifies the communication system to use wider bandwidth and the latter one brings up severe coexistence problems, as will be elaborated in Section 1.1.3.

### 1.1.1 Transmitter Linearization and Digital Predistortion

For TX, the power amplifier (PA) is the largest power consuming analog device, thereby its efficiency is an important factor that deserves to be taken into account. Fig. 1.2 shows a typical input-output relation of a power amplifier. In the low power region, the output power grows linearly with the input, but its increasing rate drops after crossing a certain point. As input power approaching the saturation point, high power efficiency can be achieved while the linearity becomes worse. Conversely, high linearity can be achieved by setting the average power large backoff from the saturation point, at the cost of reduced output power and efficiency. This is the basic linearity-efficiency contradiction of a power amplifier, which originates from the inherent physical characteristics of the transistors.

To overcome the above predicament, a few of linearization techniques, either implemented digitally or analogously, have been proposed and thoroughly studied in the past several decades [1][2]. These techniques can be categorized into several kinds, with their respective pros and cons, which are feedback, feedforward, predistortion, envelope elimination and restoration (EER), outphasing/LINC (linear amplification using nonlinear components) and others.

The use of feedback to reduce the nonlinearity of electronic devices is a practice of standard nonlinear control [3]-[5]. It runs into problems, nevertheless, when the signal propagation in the feedback loop is larger than an extra timescale that is firmly related to the bandwidth of the signal. Therefore, feedback is only applied for narrowband signal and is seldom mentioned in modern communications. The feedforward [6]-[9] is another well-established and widely used linearization technique, which once was even believed the most suitable for wideband system. It contains a correction path to offset the distortion of the signal output from the main PA, which is expensive and difficult to implement. EER [10]-

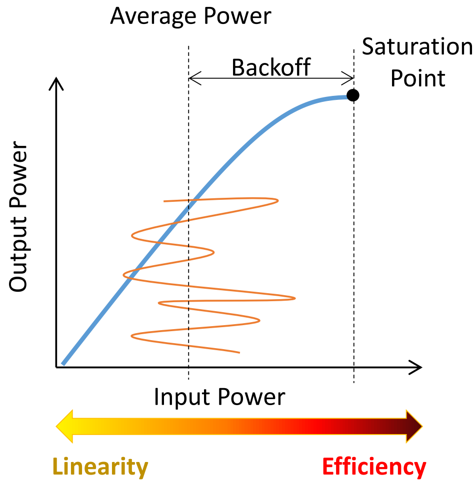


**Fig. 1.2 Demonstration of the linearity-efficiency contradiction of a PA.**

[12] and outphasing [13]-[16] also suffer the problem of limited bandwidth and high complexity, hence they are seldom found in wideband applications.

Digital predistortion (DPD) [17]-[22], which is a fruitful area of research in the past 20 years, is the most widely used linearization technique for modern wireless communications and is very likely to expand its usage for the next generation (5G). It has gradually superseded feedforward and become the most popular linearization technique for commercial wireless transmitters. The popularity of DPD is ascribed to several factors, such as simplicity for actual implementation (compared with other linearization techniques), adaptability to different signals and standards, as well as fairly good (actually the best) linearization performance. Based on such reasons, we only focus on DPD technique, which forms the main part of this thesis.

The first digital predistortion was reported by Nagata [17]. At that time, Nagata used a two-dimensional RAM (random access memory) to map the relation of I/Q signals. Because of the use of two-dimensional model to store the I/Q relation of the input and output, it contains a very large memory (20Mbits) and has a very slow convergence rate (10 seconds for signal with a data rate of 16 ksps). Later, researchers found that the distortion of the PA is only determined by the input amplitude, so that the PA can be modeled by AM-AM and AM-PM conversions. Therefore, instead of mapping the I/Q relations, the predistorter can be implemented as one look-up table (LUT) in polar form [18][19] to significantly reduce the memory size.

A boom of research on the polynomial based DPD model [23]-[27] was set off after entering into the new millennium, primarily ascribing to the evolved performance and reduced price of various digital signal processors (DSPs). Compared with LUT model, the polynomial DPD has much faster convergence rate and a variety of forms against memory effect. The shortcoming of polynomial is, however, the large number of multiplications are involved, making it unsuitable to be implemented with cheap processers and unaffordable for low-power transmitters [22].

In recent years, some researchers even start to use artificial neural network (ANN) as the DPD model [28][29], which are only limited to academic study and no practical application has been found. This is because that the PA is a system of very weak nonlinearity and memory, using very complicated model, such as ANN, cannot be afforded in practice. Again, as will be introduced in 1.1.3, current trends for the development of communication systems are enforcing designing of DPDs with low complexity, thereby this thesis attaches more attentions to practical implementation issues and tries to avoid highbrow studies.

### 1.1.2 Nonlinear Compensation for Receivers

At the RX side, the situation becomes much more complicated [30]-[34]. Fig. 1.3 demonstrated 3 cases of interference that a mobile terminal may suffer. As a receiver contains an auto gain control (AGC) to adjust the gain, which should be not too high to prevent the downstream mixer from being overdriven, but also sufficient high to prevent the receiver from failing the minimum signal-to-noise ratio (SNR) requirement. The interference caused by nonlinearity of LNA normally results from invading signals residing in other bands, with various forms.

For instance, for Case. 1 demonstrated in Fig. 1.3, when the mobile terminal is communicating with the base station (BS) D1 while locating at the edge of D1's coverage, the interfering signal from adjacent cell I is very likely much stronger than the signal from D1. The expanded spectrum of the interfering signal would overlap the wanted signal and

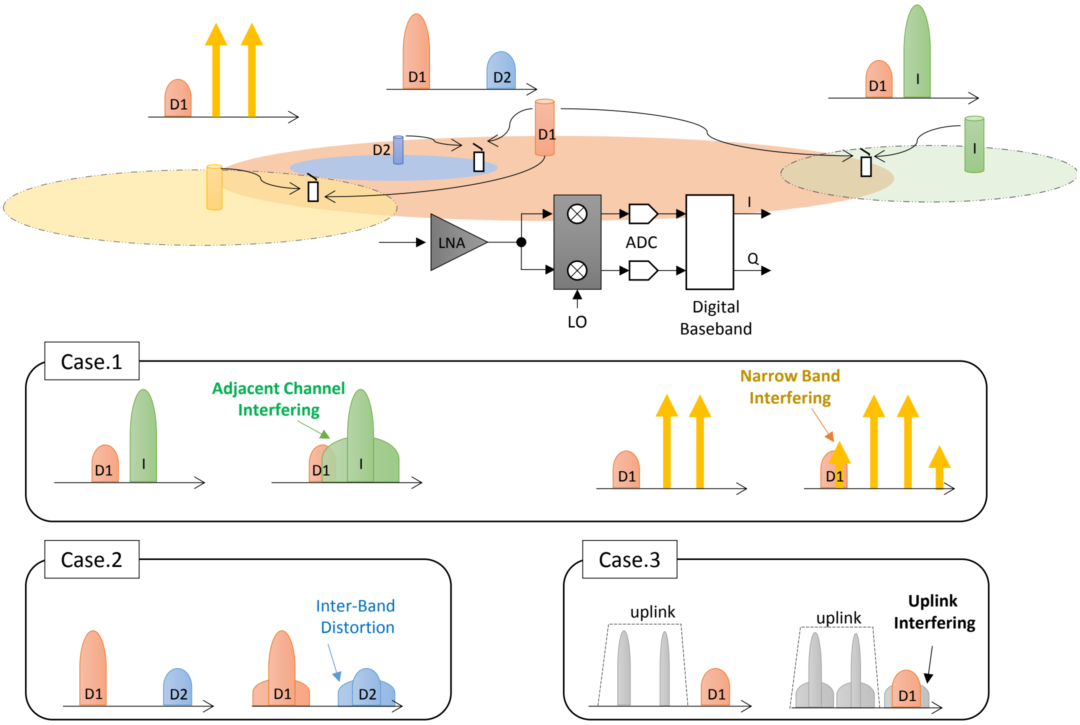
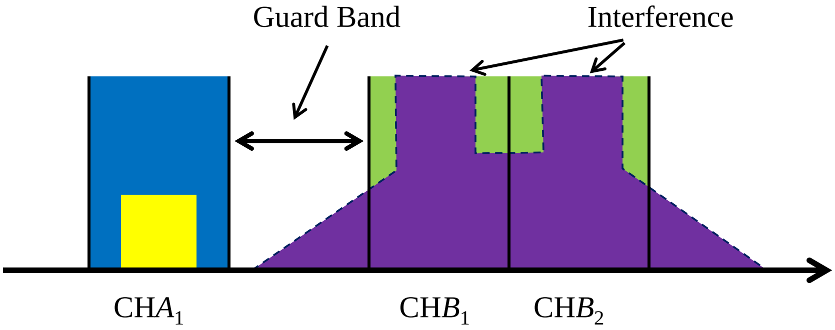
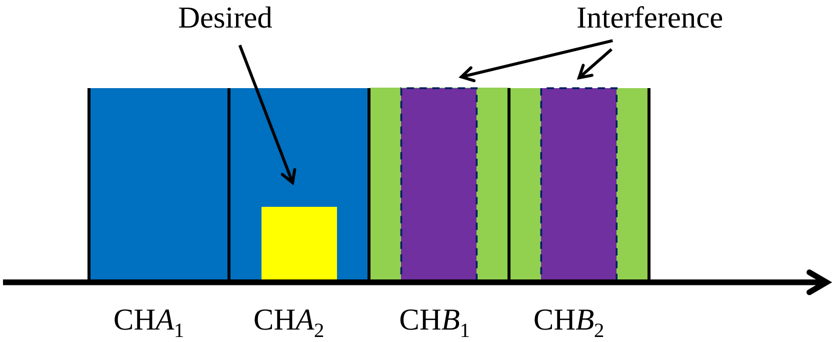


**Fig. 1.3 Demonstration of the nonlinear distorted spectrum, where the signal is an actual 20MHz LTE signal.**



**(a) A guard band is necessary to prevent interference with conventional schemes**



**(b) All bands can be effectively utilized with nonlinear compensation.**

**Fig. 1.4 Frequency band allocation**

thereby directly degrade the wanted carrier SNR. This can also be seen from narrow band interfering, for which a common example is the coexistence problem caused by GSM (global systems for mobile communications) system. As the legacy GSM cells were generally badly planed and have large coverage, such interference has become a very severe problem in many countries.

Case 2 happens in 4G LTE-Advanced system, for which carrier aggregation (CA) is utilized and the terminal is simultaneously receiving two signals from a macro BS D1 and a femto BS D2. When the two signals are received with a single front-end, cross-band intermodulation may happen, even if the two signals are located in distant bands. Case 3 happens for RX in FDD mode; when the terminal is sending the uplink signal, it can enter the receiving antenna and cause such interference.

The nonlinear interference mechanism, except for the Case 2, is result of adjacent channel leakage caused by the intermodulation distortion (IMD) of the invading signal. As the power of both the victim signal and invading signal keeps changing (which is due to the power control at the transmitter or the time varying fading channel condition) there are always some time buckets that the invading signal has much larger power than the victim signal. Thus, the nonlinear interference becomes inevitable, no matter how well the cells are planned, unless extra effort is done to prevent this happening.

The conventional way to avoid nonlinear interference is to set a blank band to separate the desired signal and potential interference, which is demonstrated in Fig. 1.4(a). Suppose that CH$A_1$ is the channel used by the operator/terminal $A$, while CH$B_1$ and CH$B_2$ are used by operator $B$ or used for uplink. A guard band is necessary between them to prevent intermodulation interference in practical situations. Conversely, if the front-end of the receiver is well linearized in a large power range, the occasion of such interference can be avoided, hence the guard band becomes unnecessary, allowing effective utilization of precious spectrum resource. For instance, $A$ can use the protection band as a new channel CH$A_2$ shown in Fig. 1.4(b).

Many techniques have been proposed to combat the nonlinear distortion of receiver front-ends. In general, they can be categorized into two groups. One is the feed-forward technique [35]-[37], in which the IMD cancelling signal is generated in an auxiliary path and subtracted from the main path to cancel the nonlinear distortion. Another one is the post nonlinear compensation (PNLC) technique, which linearizes the distorted signals by emulating the inverse characteristics of the nonlinearity [38]-[42]. One of the demerits of the former scheme is that the cancellation can only be performed with analog circuit, lacking of flexibility for reconfiguration and versatility for different standards. Furthermore, because of the difficulty and complexity for implementing the analog signal processing, it is desirable to shift such tasks into digital domain. With the development of high-speed low-power ADCs (analog-to-digital converters), the idea once seemed impractical has become a promising solution. Together with the recent IC (integrated circuit) technologies, advanced digital signal processing techniques can fulfill such a goal.

This thesis follows the later strategy (PNLC) and tries to develop compensator exploiting baseband signal processing. The nonlinear characteristic of the front-end is generally unknown, thus blind identification algorithm needs to be applied to achieve real-time adaptive compensation.

### 1.1.3 Challenges Raised by the Current Trend of Digital Wireless Communications

#### *1.1.3.1 Increasing Bandwidth*

Digital wireless communications have experienced explosive demands for higher date rate and have kept evolving to offer faster transmission speed [43]. Fig. 1.5 has summarized the downlink speed for some representative mobile systems that have been widely used in their respective eras, from the initial digital communication 2G to the recent 4G LTE-Advanced. One can see the dramatic increasing of supported data rate, especially in the last decade.

The need to support the mobile data traffic explosion is unquestionably the main driving forcing behind the development of digital wireless communication. The mobile data traffic has grown more than 24-fold between 2010 and 2015, and is foreseen to be 500-fold between

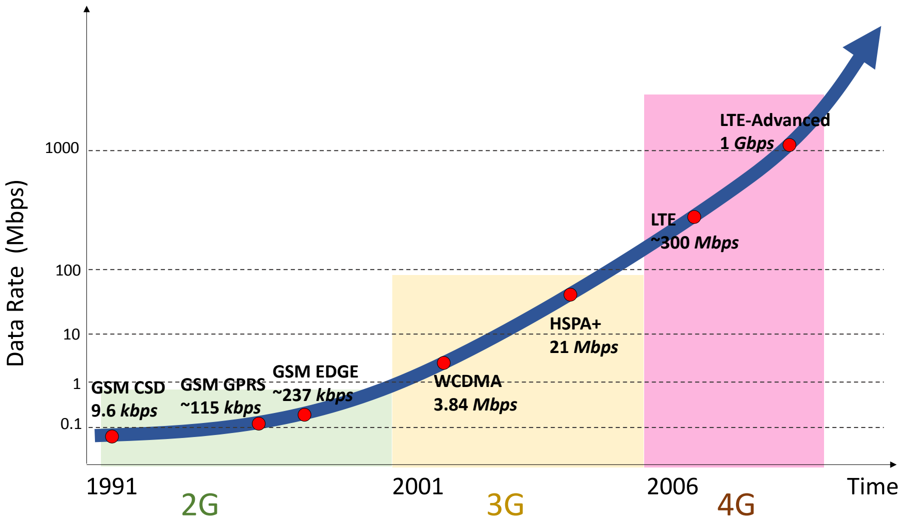


**Fig. 1.5 Demonstration of the nonlinear distorted spectrum, where the signal is an actual 20MHz LTE signal.**

2010 and 2020. To handle this, the next generation 5G [44]-[46] is believed to be faster than 4G by from 100 to 1000 times.

With the ever increasing demands for higher data rate, the bandwidth of contemporary wireless communication systems also have had to continuously increase to improve the throughput. The direct consequence of increased bandwidth is the higher speed requirements imposed on processors and data converters. Higher the speed (clock frequency) the larger power being dissipated by the digital circuits. As the digital circuits also take up a large portion in the modern wireless BSs, the power consumed by digital circuits can easily surpass the power consumed by the analog circuits.

The minimum speed of the data converters is generally 5 times (or more) of the Nyquist rate of the signal to be processed, in order to include the intermodulation component caused by nonlinearity. The communication system on active duty, e.g. the 4G LTE-Advanced has a maximum contiguous signal bandwidth of 100 MHz. The Nyquist rate for such signal, depending on the transmitter and receiver, will be 1 GHz for superheterodyne structure or 500 MHz for zero-IF structure [47]. For either case, such off-the-shelf high-speed data converters, especially for the analog-to-digital converters (ADCs), are quite expensive, which may even surpass the price of the high power components, like the PA. Therefore, utilizing hardware with reasonable price to perform nonlinear compensation is a challenge raised by such high-speed systems.

On the other hand, a predicament that mobile operators are being faced with is that they do not have enough contiguous usable bandwidth, thus transmitting two signals in different bands as opposed to one band to form a wideband signal leads to the usage of carrier aggregation (CA) [48]-[51], which is a key technology for LTE-Advanced. This pushes the study on concurrent dual-band transmitter, as it reduces the hardware expense by using a single

PA. However, the distortion caused by the basic nonlinear characteristics of the PA causes an urgent need for the study on dual-band DPD, which has an absolute different distortion mechanism compare to conventional DPD.

#### *1.1.3.2 Coexistence of Multiple Standards*

The development of transceivers for modern mobile communication systems is subject to the plethora of standards have been developed to date [52][53]. To support universal global communications anytime and anywhere, one cellphone should simultaneously support at least six different standards: 2G or 2.5G (GSM, GPRS, EDGE), 3G (WCDMA, CDMA2000 or TD-SCDMA), 4G LTE, WLAN (802.11 series), Bluetooth and Navigation (GPS). If each communication mode is implemented with independent hardware, the BOM (bills of materials) cost would be expensive and the form factor (size) would be large since multiple sets of circuits in parallel are needed. Considering that multiple communication applications are not used at the same time, it would be cost efficient to reuse the hardware and dynamically reconfigure its communication mode according to the user's command.

Together with the constant emergency of new technologies and diversified functions of user-end, as well as the above multiple standards, the radio environment is successively getting more and more complex. In the context of multi-mode mobile terminals for which multiple radio systems must be simultaneously supported, new coexistence challenges have been created. Generally, the coexistence issues arise due to the receiver performance is negatively affected by the signal from another transmitter, as already presented in Section 1.1.2. The blocking mechanisms of the receiver are quite complicated, but the issue that with most concern is the direct rise of the desired signal’s noise floor caused by the interferer’s out-of-band radiation, as portrayed in Fig. 1.3. Normally, a sound surface wave filter is used to block the interferer, which, however, has limited attenuation so that a nonlinear compensation is applied to improve the linearity of the receiver front-end.

As market pressures demand the further reduction in the expense and size of the off-chip components, auxiliary blocks that once facilitated receiver front-end linearization, such as filters and feedforward, will be superseded by on-chip solutions. This tendency is to miniaturize the devices and transfer more functions from off-chip devices to digitally-controlled circuits on a single chip.

#### *1.1.3.3 Paradigm Shift from Homogeneous to Heterogeneous Networks*

Utilizing higher-order modulation and larger bandwidth to increase the throughput will meet the limit, both technologically and economically. Along with the motivation of increased capacity and reduction in cost and environmental problem, heterogeneous network [54]-[56] (a hybrid network compassing cells of various sizes), which includes advanced transmitter and receiver technologies, is being developed to fulfil this goal. Meanwhile, building power-efficient green radio architecture is also a key consideration for building

**Table 1-1 Base station power consumption at maximum load for different cells, where the transmitters are examples for a commercial LTE system.**

| | TX Power (W) | PA Peak Power (W) | Total Power (W) |
|---|---|---|---|
| Macro | 20 | 130 | 225 |
| Micro | 6.3 | 28 | 72 |
| Pico | 0.13 | 1.9 | 7.3 |
| Femto | 0.05 | 1.1 | 5.2 |

nowadays networks [57][58].

A vital feature of heterogeneous network is the coexistence of macro, micro, pico and femto cells. Table 1-1 shows the power data from an LTE system. One can see that the macro BS has an average transmission power of 20 W and uses a PA with 130 W peak power. The average power is much smaller than the maximum output capability the PA because of the large backoff set to the input signal. Along with the power of the digital portion, the macro BS consumes 225 W (without considering cooling equipment), where the power of the PA takes up over half of the total power consumption.

The PAs for micro BSs account for less than 30% of the overall power consumption, and this proportion is even lower for pico and femto. In such small power BSs, the efficiency requirement for PAs is much relaxed, so that it can be allowed to operate with larger backoff.

As market pressures demand the further reduction in the expense and size of the cell, there are additional opportunities for further integration of the digital radio front-end technology and the baseband processing executed by system-on-chip (SoC) solutions, as well as portions of the radio front-end with the analog RF circuitry. In a typical macro BS, the functional architectures are generally divided into three parts: control processing, baseband processing and front-end (a more detailed description about the front-end will be presented in the next session), which are formed by a number of integrated circuits and several boards partitioned along functional boundaries. As small cell has become the central topic for heterogeneous network, the demand posed here is that they should be much smaller and consume less power and hence, aggregating less circuits on a single board.

Therefore, the DPD once implemented on single chip, which can be afforded by the macro BS, will be no longer practical for small BSs. Then a question may arise that "*whether the DPD is still necessary for small BSs*?"

**The answer is a definite *yes*.**

First and foremost, to build a front-end with small form factor, the performance requirement that imposed on the RF systems can be much relaxed with the facilitation of digital compensation techniques. For instance, the front-ends for small cell always use sound acoustic wave filters and duplexers, which are much smaller and cheaper than ceramic or cavity devices that are used in high-power transmitters. They introduce an extent of power loss in several decibels, and thus the PA needs to work with higher power, pushing them into nonlinear operation region. This justifies the use of linearization techniques in

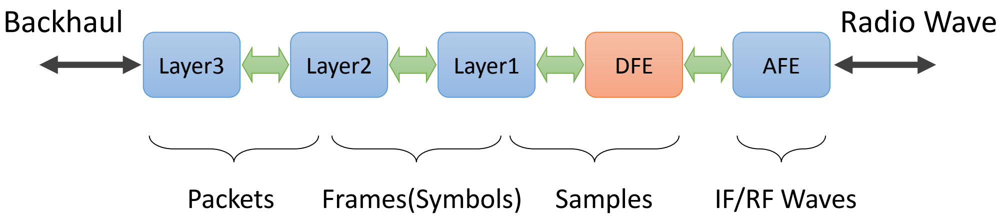


**Fig. 1.6 Demonstration of functional partition for a typical base station, where IF stands for intermediate frequency.**

order to ensure the transmission requirement set by the communication standard.

To give small form factor, moreover, small BSs are deployed with enclosures that exclude air cooling, and hence suppressing the power dissipation becomes very important. Linearization techniques like the DPD can help to significantly lower the overall power consumption, enabling the choices of cheaper PAs and less off-chip components, which yields less cost. It is also worth to mention that the DPD encloses a receiving path that is composed of down converters and ADCs, which adds extra BOM. Therefore, the DPD should be made with minimized cost and complexity, making it a worthwhile deal for small cells.

## 1.2 ABOUT THIS THESIS

### 1.2.1 Technical Scope and Features

This thesis studies the DPD for transmitters and PNLC for receivers, both of which are based on digital signal processing in baseband. Each of these two techniques, especially the DPD, serves as an important role for wireless communication, but they are not categorized to a necessary layer of the communication system and instead they commonly seen as parts of digital front-end (DFE). Fig. 1.6 demonstrates the position of DFE in a typical base station. In the transmission phase, the DFE inherits samples from layer 1, which is known as physical layer, and feeds the processed samples to analog front-end (AFE) through data converter such as a digital-to-analog converter (DAC). For receiving phase, the direction of above flow is just reversed.

The DFE cares less about the modulation details of the signals, as which seen from the DFE are only streaming and random samples. A typical DFE contains the functions of resampling (interpolation and down sampling), filtering, digital compensation, crest factor reduction (CFR), frequency shifting and etc., but which may also differ depending on the structure of the front-ends used in transmitter/receiver. The commonly used front-end structures are digital-IF and zero-IF, as diagrammatically depicted in Fig. 1.7 and Fig. 1.8. In digital-IF structure transmitter, for example, the baseband signal is converted to IF using digital up converter (DUC), which is composed of interpolator and frequency shifter, and then transferred to analog signal with a DAC. The receiver has a bi-univocal architecture that has the reversed direction as the transmitter. The zero-IF, on the other hand, has a

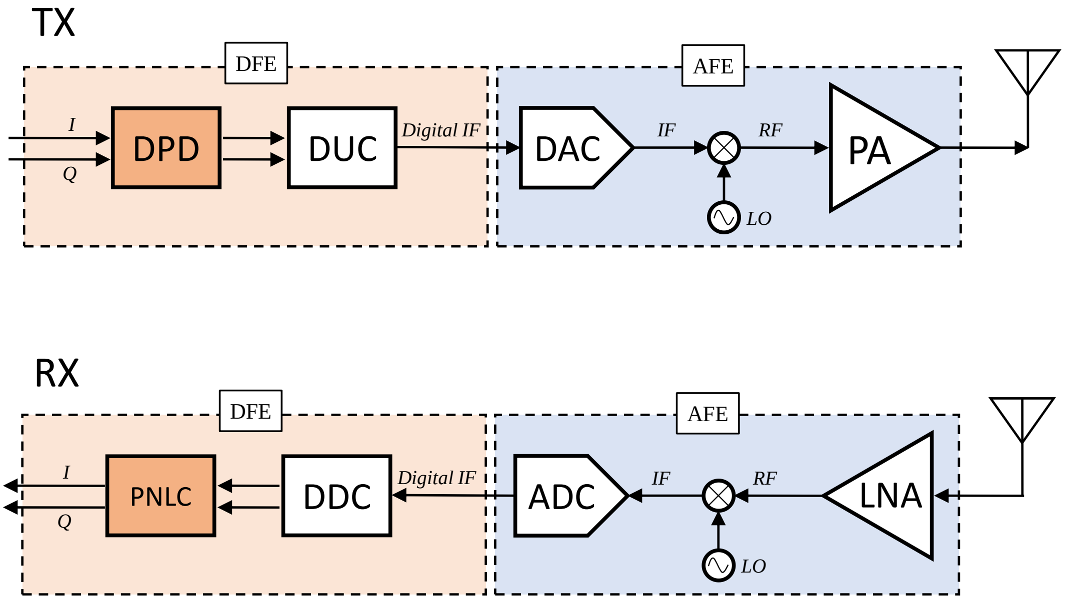


**Fig. 1.7 Simplified diagrams for the front-ends for digital-IF structure transmitter and receiver.**

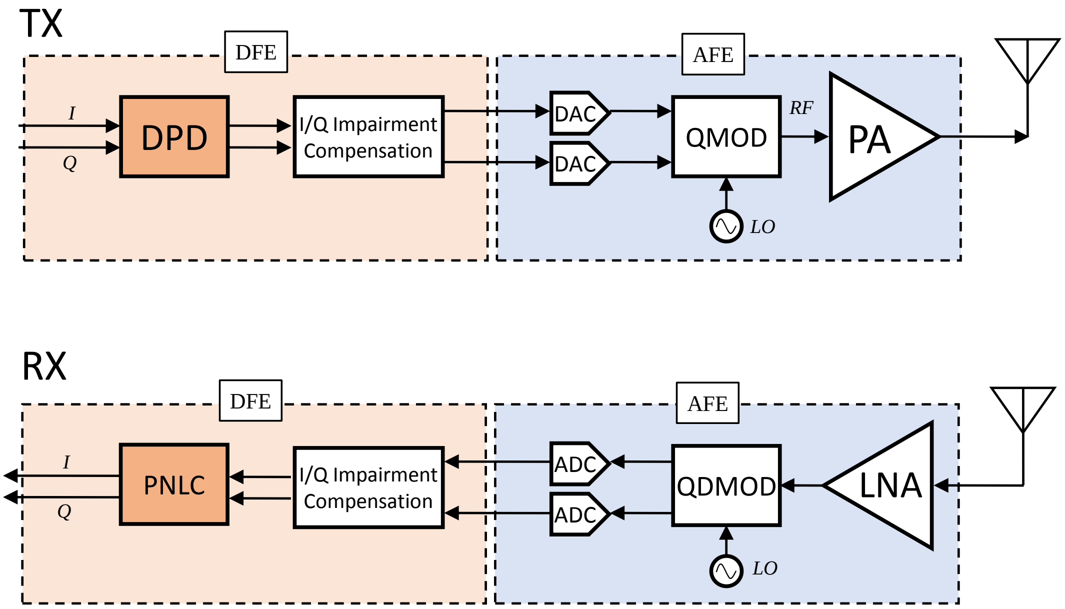


**Fig. 1.8 Simplified diagrams for the front-ends for zero-IF structure transmitter and receiver.**

direct conversion between baseband and RF, which is realized by the analog quadrature modulator (QMOD) and quadrature demodulator (QDMOD). Zero-IF reduces the basic circuit complexity by performing only one time of frequency conversion and lowers the Nyquist rate requirement for the data converters, but which in turn increases a path to convert the respective I and Q signals and also causes problems such as I/Q imbalance and DC offset. Therefore, I/Q impairment compensation becomes crucial to ensure the zero-IF operation.

The DPD and PNLC locate at the beginning and end of the DFEs, so that they functionally serve as pre-compensation and post-compensation of the nonlinear AFE, respectively. The input for DPD is intact baseband signal or the signal after CFR, while the input for PNLC is the distorted signal. For either transmitter or receiver, the nonlinear mechanism is similar but the strategy taken to perform compensation is different. At the transmitter, as both the intact signal and the distorted signal (through a feedback receiving path) are available, the nonlinearity can be accurately modeled and thus the compensation can be performed with high precision. For receiver, however, only the distorted signal is known, so that the PNLC represents a blind nonlinear modeling problem and which is generally less accurate than its transmitter counterpart [30]. Moreover, in many situations, the signal PNLC undertaken contains both the desired signal and the strong interfering signal, so actually it is to compensate the distortion of the interfering signal and leave the desired signal.

Instead of studying nonlinear models and identification algorithms, this thesis attaches importance to technological realizability of the DPD and PNLC, and proposes new techniques that can fulfil the aggressive designing goals for DPD and PNLC with reasonable hardware complexity. This includes, for instance, FPGA implementation of DPD for actual commercial transmitters, proposing of DPD technique that breaks the bottleneck set by the ADC, novel blind PNLC/DPD technique used for concurrent dual-band receiver/transmitter linearization, and so on. The test signals used in this thesis are actual standardized signals used in mobile communication systems, which are GSM, WCDMA (Wideband Code Division Multiple Access) and LTE/LTE-Advanced, ranging from antique 2G to latest 4G.

The work here is carried out using three approaches: computer simulations; experimental methods using instruments; practical implementation with FPGA. Simulations are taken place using MATLAB, which is a powerful tool for matrix calculation and numerical computation. Measurements using instruments are based on SG-SA (SG *abbrv. for* vector signal generator; SA *abbrv. for* vector signal analyzer) test bed, which provides quick assessment to the suitability of the model and algorithm to actual nonlinear devices. The implementation on FPGA considers practical engineering issues, in order to involve the proposed techniques in more realistic situation.

### 1.2.2 Original Contributions

This thesis makes several original contributions to nonlinear compensation for wireless transmitter (DPD) and receiver (PNLC), and they are summarized below:

1) A recursive method for inverting the general nonlinear function is proposed and successfully introduced to DPD for inverting the PA model. The recursive inverse is expanded to form a forward structure, hence it can be fully pipelined allowing high-

speed implementation. A method for finding the inverse of a given baseband polynomial is proposed, which is much faster than numerical inverting methods. They are contained in Chapter 3 and appears on

Yuelin Ma, Songbai He, Yoshihiko Akaiwa and Yasushi Yamao "An open-loop digital predistorter based on memory polynomial inverses for linearization of RF power amplifier," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 21, no. 5, pp. 589–595, Sep. 2011.

Yuelin Ma and Yasushi Yamao, "A Method for Inverting Monotonous Continuous Nonlinear Function," 2014 TriSAI Workshop, pp. 1-5. Beijing, China, Sep. 2014

Yuelin Ma, Yasushi Yamao and Yoshihiko Akaiwa, "An algorithm for obtaining the inverse for a given polynomial in baseband," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E96-A, no. 3, pp. 675–683, Mar. 2013.

Yuelin Ma, Yasushi Yamao and Yoshihiko Akaiwa, "Fast Baseband Polynomial Inverse Algorithm for Nonlinear System Compensation," IEEE Proceeding of Vehicular Technology Conference 2012-Spring, 2F-1, Yokohama, Japan, May 2012.

2) A real-time DPD with low complexity and fast convergence rate is implemented in a FPGA to linearize commercial Doherty amplifier used in radio remote head and repeater systems. This DPD has strong robustness against fast variation of the signal status and can achieve very good linearization performance. This work is presented in Chapter 5 and has been published on

Yuelin Ma, Yasushi Yamao, Yoshihiko Akaiwa and Chunlei Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol.61, no.11, pp.3961-3973, Nov. 2013.

3) The ADC is a very expensive component, which hurdles the application of DPD for ultra wideband signals. The author proposed a spectral extrapolation based DPD that can handle signals with much wider bandwidth than the ADC's sampling rate. Experiments using this technique with a signal of 160 MHz bandwidth were successfully carried out. This is, to the author's best knowledge, the DPD with the widest bandwidth in the world. This is the main part of Chapter 6, and was reported in

Yuelin Ma, Yasushi Yamao, Yoshihiko Akaiwa, Koji Ishibashi, "Wideband digital predistortion using spectral extrapolation of band-limited feedback signal", *IEEE Transactions on Circuits and Systems I*, vol.61, no.7, pp.2088-2097, Jul. 2014.

4) We also turn to the issue of interference caused by nonlinear receiver front-end, where a real-time blind PNLC is proposed and evaluated with experiments built on FPGA. The contents are shown in Chapter 7 and have also appeared in

Yuelin Ma, Yasushi Yamao and Yoshihiko Akaiwa, "Blind Nonlinear Compensation Technique for RF Receiver Front-End," 43rd European Microwave Conference, pp. 1527-1530. Nuremburg, Germany, Oct. 2013.

5) It is extended to the receiver nonlinear issue in previous chapter to concurrent dual-band receiver. The nonlinearity is compensated using a bi-variant polynomial model, which is validated with instrumental experiments. This content is presented in Chapter 8 has been published on 1 journal paper:

Yuelin Ma, Yasushi Yamao, Koji Ishibashi and Yoshihiko Akaiwa, "Adaptive compensation of inter-band modulation distortion for tunable concurrent dual-band receivers," *IEEE Transactions on Microwave Theory and Techniques*, vol.61, no.12, pp.4209-4219, Dec. 2013.

6) A spectra-folding feedback path architecture for concurrent dual-band power amplifier predistortion, to simplify the feedback path and significantly reduce the cost. It has the halved BOM cost and non-sacrificed performance, thus it has the potential to replace the conventional dual-band DPD architecture. This content is presented in Chapter 9.

### 1.2.3 Outlines

This thesis composes 10 Chapters, which are outlined in Fig. 1.9 and the organization is described as follows:

- **Chapter 1** is the introduction that contains the background and motivation of this thesis.
- **Chapter 2** gives a brief introduction of the basics about digitally modulated signals and nonlinear amplifiers to provide a useful technique background knowledge.
- **Chapter 3** sets up the mathematical formulations for the nonlinear models, including mathematical expression and linear regressive architecture of the polynomial model. The methods for orthogonalization and pruning of the model are also presented. It is followed by the presentation of a general mathematical method for inverting the linear model as well as the method for obtaining the inverse of a given polynomial.
- **Chapter 4** describes practical issues regarding instrumental experiment and FPGA implementation. Some issues that, such as I/Q impairment and gain/delay mismatch, must to be resolved in practice are discussed.
- **Chapter 5** presents an FPGA implementation of a DPD with low complexity and high convergence speed for multi-carrier GSM (MC-GSM).
- **Chapter 6** considers the limitation set by ADC and proposed a novel DPD technique addressing very wideband signal. A DPD with the widest bandwidth that ever been reported was realized using the proposed technique.
- **Chapter 7** turns to the issue of interference caused by nonlinear receiver front-end, where a real-time blind PNLC is proposed and evaluated with experiments built on FPGA.
- **Chapter 8** extends to the receiver nonlinear issue in previous chapter to concurrent dual-band receiver. The nonlinearity is compensated using a bi-variant polynomial model, which is validated with instrumental experiments.
- **Chapter 9** proposes a spectra-folding DPD architecture for concurrent dual-band power amplifier linearization, to simplify the feedback path and significantly reduce the cost.
- **Chapter 10** Conclusions are finally drawn here.

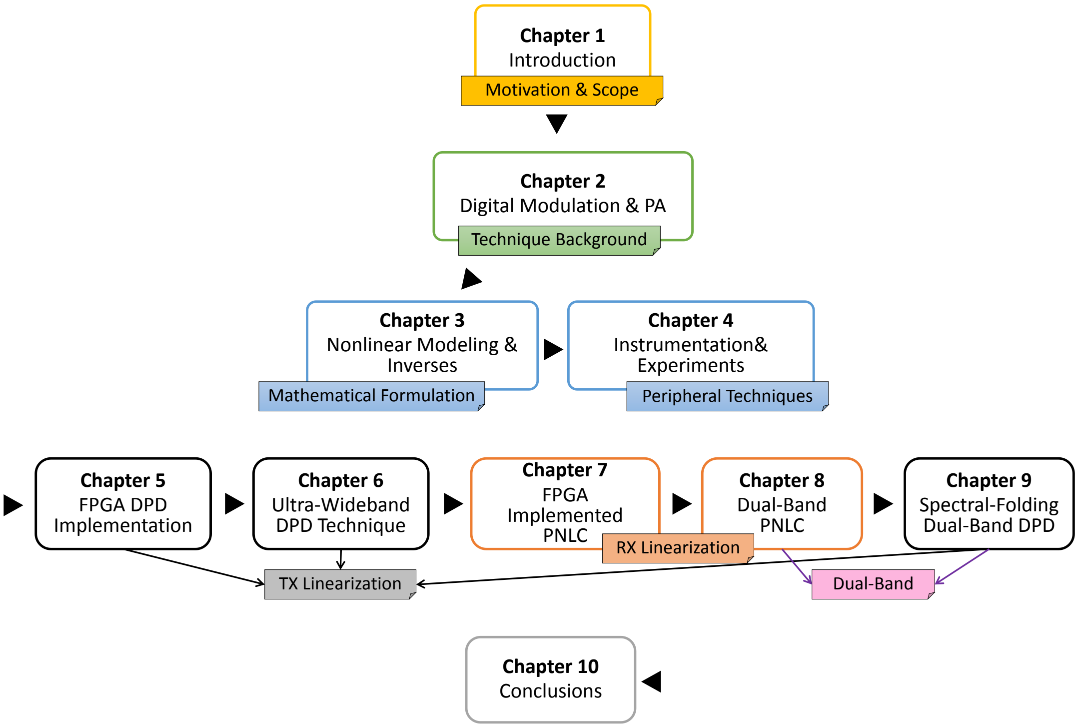


**Fig. 1.9 Outline of this thesis.**

# Chapter 2.

# Digitally Modulated Signals and Nonlinear Power Amplifiers

## 2.1 DIGITALLY MODULATED SIGNALS

As nonlinear compensations considered in this thesis are taken place in baseband, the nonlinear characteristics are a reflection of the actual devices to baseband. It is natural to think that the property of the signals can affect or even change the equivalent baseband response, since we use baseband model to analyze and characterize the devices. This section provides a brief overview of digitally modulated signals concerning the signal properties rather than modulation mechanisms. In addition, since WCDMA (wideband code division multiple access) and LTE (long term evolution) signals are used for measurements in this thesis, we also present a brief introduction for CDMA (code division multiple access) and OFDM (orthogonal frequency division multiplex), which serve as the key technique for WCDMA and LTE, in Section 2.1.2, and the large PAPR (peak-to-average power ration) problem in Section 2.1.3. Furthermore, the multi-carrier GSM (MC-GSM), as a thorny issue and trouble maker in actual DPD applications, is also introduced. The figure-of-merit (FoM) for benchmarking the amount of distortion of the test signal, including the well-used ACLR (adjacent channel leakage ratio), EVM (error vector magnitude), NMSE (normalized mean square error), is introduced in Section 2.1.5.

### 2.1.1 Baseband and Passband Expressions

The digitally modulated signals in modern wireless communication systems are generally two-dimensional variation in amplitude and phase, hence the baseband signal has a general form of [59]

$$x(t) = A(t)e^{j\theta(t)}. \tag{2.1}$$

According to the theorem of information entropy, baseband signal should be a random time sequence when the amount of information of the baseband signal is maximal. Accordingly, the signal in process, which is generated from baseband signal through interpolation, is a band limited complex random signal.

It is obvious that the processing of baseband signal is either on polar basis (amplitude and phase) or quadrature basis (I and Q). In despite of any structure of transmitter is applied, the baseband signal is related to the corresponding RF (radio frequency) or passband signal $\tilde{x}(t)$ through

$$\tilde{x}(t) = \Re\left\{A(t)e^{j\theta(t)}e^{j2\pi f_c t}\right\}, \tag{2.2}$$

which can be written as

$$\tilde{x}(t) = A(t)\cos\left(\omega_c t + \theta(\mathrm{t})\right), \tag{2.3}$$

where $\omega_c$ is the center frequency in radian. This function indicates that the digitally modulated signal can be seen as a simultaneous amplitude- and phase-modulated signal, and this expression will be repeatedly used in this thesis.

### 2.1.2 CDMA and OFDM

CDMA and OFDM are widely used in 3G and 4G mobile system [60]-[63] as well as WLAN (wireless local access network) to offer key advantages such as flexible allocation of resources and high spectrum utilization.

CDMA is a spread-spectrum multiple access technique, which spreads the bandwidth of the data uniformly for the same transmitted power. The transmitted signals are multiplied through XOR (exclusive OR) with different spreading codes with much faster rate. The spread signals are combined by addition to form the CDMA signal. The spreading codes are mutually orthogonal, so that the transmitted signals can be identified at the receiver.

OFDM signal is a combination of a number of sub-carriers, which are narrow band signals. The sub-carrier frequencies are chosen so that the sub-carriers are orthogonal to each other, meaning that cross-talk between the sub-channels is eliminated and inter-carrier guard bands are not required.

From the above descriptions, it can be seen that CDMA is a combination of signals orthogonal in time domain, while OFDM is a superposition of signals orthogonal in frequency domain. Recalling the basic nature of duality of time domain signal and frequency domain signal, CDMA as opposed to OFDM have somewhat similarities. As they are combination of a number of random signals, both of them approximates Gaussian distribution according to central limit theorem. This causes large PAPR problem, as will be shown soon later.

### 2.1.3 Large PAPR Problem and Crest Factor Reduction

As CDMA and OFDM signals are essentially a sum of multiple signals/codes in either time or frequency domain, their probability density functions (PDFs) for real and imaginary part (I and Q) approximate Gaussian distribution. It is well known that the amplitude of a 2-dimensional Gaussian signal follows Rayleigh distribution, for which the PDF has a long tail. This results in large PAPR, which poses strict demands on the dynamic range of data converters and especially limits efficient operation of the power amplifiers (PAs). The probability distribution for actual WCDMA and LTE signals are plotted in Fig. 2.1, which are in

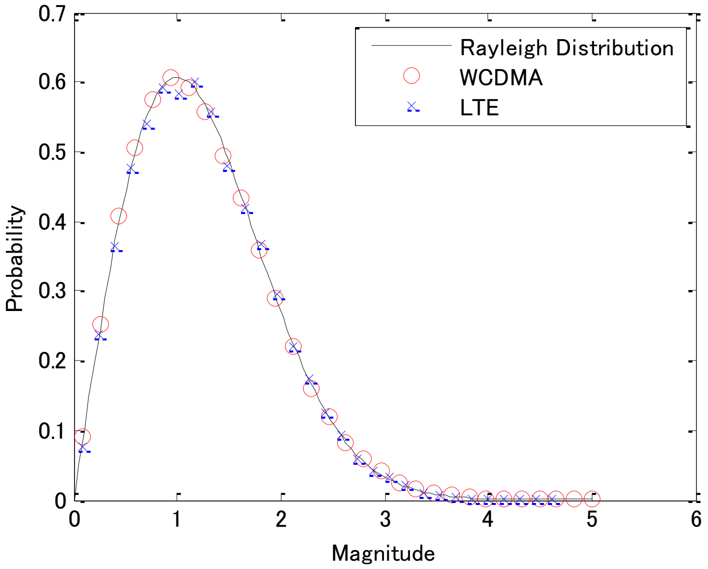


**Fig. 2.1 Probability distribution for WCDMA and LTE signals, which firmly fit the Rayleigh distribution**

comparison with a normalized Rayleigh distribution. The WCDMA signal used here complies with the 3GPP test model 3 with 64 multiplexed channels, and the LTE signal is based on FDD (frequency division duplex) test model 3.3 with 5 MHz bandwidth. It can be seen that both of WCDMA and LTE firmly fit the standard Rayleigh distribution PDF.

Large PAPR also hampers the performance boost that can be achieved by the DPD. To appreciate this statement, it is beneficial to first review some basic concepts relating the PA behavior, which is demonstrated in Fig. 2.2 to provide an intuitive vision. The $P_a(\,)$ and $P_p(\,)$ used here is to denote the average power and peak power of the designated signal. The basic nonlinearity of the PA can be described as a suppressive behavior, as shown in the Figure, which is more specifically a tendency of reduced gain as input power getting higher. The input backoff (IBO) or output backoff (OBO) is the difference between input/output saturation power and input/output average power. Normally, we keep driving the signal away from the saturation region to avoid severe clipping distortion, so that we always have

$$\begin{aligned} \text{IBO} &\geq \text{PAPR}(v) \\ \text{OBO} &\geq \text{PAPR}(y). \end{aligned} \tag{2.4}$$

Moreover, to guarantee the PA invertible seen from the baseband, DPD only work in the area up to the PA's saturation point, hence it sets a *forbidden region* that the signal cannot be transgressed, shown as the shaded area in Fig. 2.3. Consequently, the conditions set in (2.4) need to be strictly ensured when the DPD is applied.

By taking a glance of Fig. 2.3, one can observe that the DPD has an expanding behavior as opposed to the suppressing behavior of the PA. Seeing from the bottom of Fig. 2.3, which shows the input signal $x$ is enlarged by the DPD resulting in larger Pp($v$) than the that in

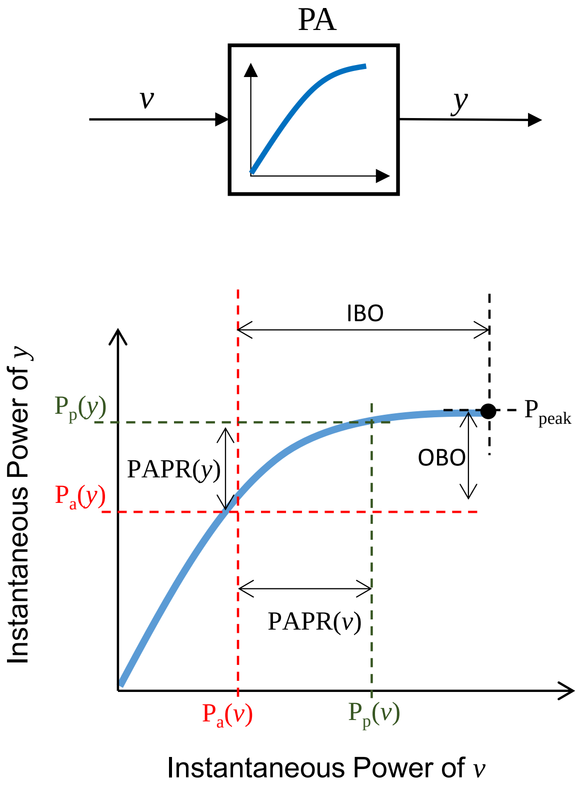


**Fig. 2.2 Input-output relation of the PA.**

Fig.2.2. Therefore, it is not surprising that $v$ has a larger PAPR than $x$, implying that larger backoff is required after the DPD is taken effect in order to prevent the DPD from the *forbidden region*. This is absolutely adverse to power efficiency, and also raises a dilemma for DPD that if the same amount of backoff is set for the PA without DPD, similar linearity may also be achieved.

After the observation of the negative effect of large PAPR, reducing the PAPR hence becomes important for boosting the PA efficiency and DPD effectiveness. There are a myriad of literatures considering the crest factor[1] reduction (CFR) techniques [65]-[67], most of which, however, are too complicated and lack of the implication in practical applications. In addition, for most CFR techniques, the resulting CCDF (Complementary Cumulative Distribution Function) has a parabolic curve, for which the actual peak is intractable. As one should keep the signal away from the *forbidden region* as explained before, the peak need to be strictly limited into a predefined range. Thus, it is nature to consider that the CFR should yield a signal with desired peak.

Peak cancellation, which has the merits such as low complexity and tractable resulting peaks, has proved itself the most suitable CFR technique for practical application. In recent

---

[1] Crest factor is defined as the ratio of peak amplitude to average amplitude, so it has a square root relation to PAPR

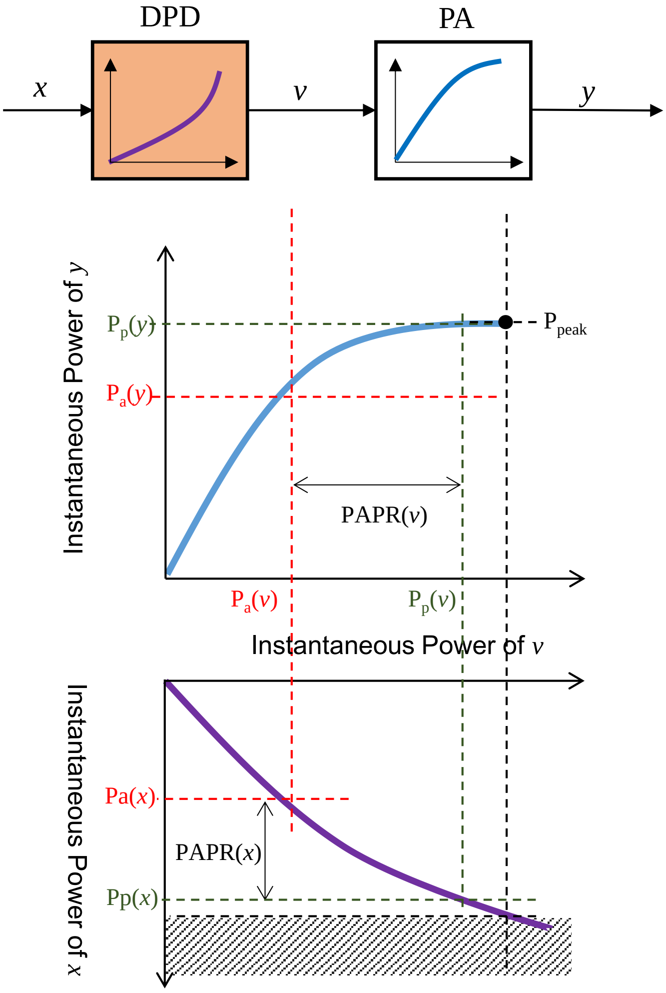


**Fig. 2.3 Input-output relation of the DPD and PA.**

years, some off-the-shelf CFR products utilizing peak cancellation have also become available [68][69]. The peak cancellation reduces the PAPR by subtracting a cancelling pulse from the peaks that surpass the predefined threshold. An extent of distortion in terms of increased EVM (Error Vector Magnitude) inevitably occurs, but it will be paid off with reduced PAPR and boosted power efficiency. Considering that, the transmitter can tolerate a specific extent of EVM deterioration, peak cancellation causes no problem as it has negligible effect compare to wireless channel conditions such as fading and noise.

As for most PA the power supply and bias can be considered as constant, the signal envelop would also affect the dilivered output power, as demonstrated in Fig. 2.4. Here we simply assume the constant power supply define the bound of the peak output power, such that the shaded area can be seen as the wasted power. The DPD, as already discussed before, expands the signal amplitude, resulting a streched long tail for the amplitude PDF, see right upper of Fig.2.4. The CFR, on the other hand, reduces high peaks to a desired threshold, so that smaller backoff is allowed, therein larger portion of the signal is driven into the high

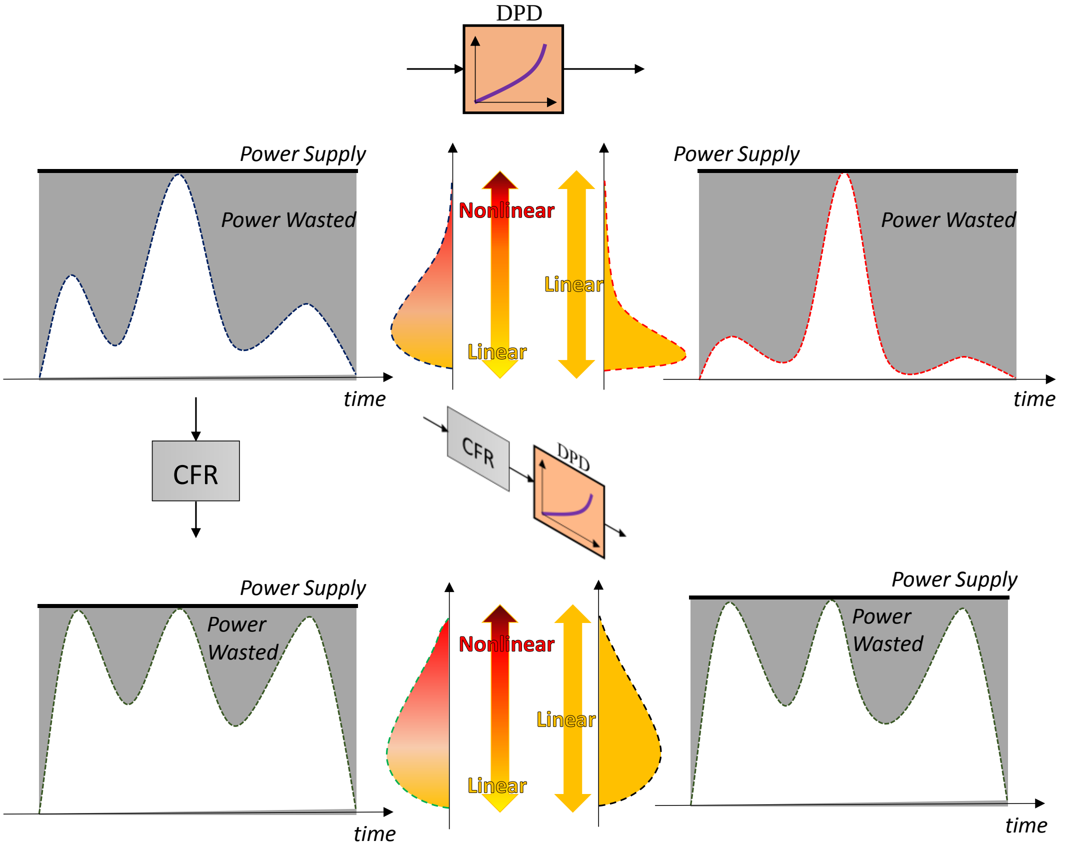


**Fig. 2.4 Demonstration of the achievable power efficiency for CFR, DPD and CFR+DPD.**

efficiency (large nonlinear distortion) region. Only CFR is combined with DPD, can both high efficiency and high linearity be achieved.

## 2.1.4 Multi-Carrier GSM

Albeit a legacy communication system, GSM, including EDGE (Enhanced Data Rates for GSM Evolution) and GPRS (General Packet Radio Service), can still find its usage in many countries even until now. Global GSM subscribers have reached a large base of 6.9 billion, and which is expected to be 8.5 billion by the end of 2016 [70]. Although new generation of communication systems are being developed world wide, the GSM still represents a significance that will continue to exist. However, GSM is not friendly to the other new systems and actually causes severe coexistence problem as the one shown in Chapter 1.1.2. Therefore, studying the GSM, especially the problems existing in practical application, is still meaningful. We bring GSM up here is not because of its advances in technique, but in fact it poses severe challenges for both transmitter and receiver design, especially in wideband implementations.

GSM forms a narrow-band digital communication system, which originally can provide a data rate no more than several hundreds of kbit/s and has a channel bandwidth of less than 200 kHz. The narrower the bandwidth, the more difficult the design of IF and digital-to-analog converter (DAC), thus GSM actually represents a tougher object than the state-of-the-art communication signal, such as LTE, for analog front-end (AFE) design.

With the advent of wideband RF technique, multi-carrier GSM (MC-GSM) has been widely used to meet the demand on higher throughput. In addition, pushed by aggressive pursuit for lower cost and enabled by advances in integrated circuit and multi-carrier PA technology, the current GSM system (5th generation GSM) forms the multi-carrier IF (intermediate frequency) signal with digital processing and generates the analog signal with a DAC. Thus, no combiner is needed and the power consumption is greatly cut down.

Although the GSM carrier has only a small occupied bandwidth, when multiple of them are aggregated together, the used bandwidth can still be large. Like the CDMA and OFDM, MC-GSM can also be seen as an addition of multiple random signals; hence, it also has a large PAPR. In practice, nevertheless, depending on the available bandwidth of the operator and other factors, the number of GSM carriers ranges from several to several tens. This implies that the MC-GSM is unlikely to be a Gaussian distributed signal, so that the assumptions made to Gaussian signal that has been applied to WCDMA and LTE is no longer applicable to it.

MC-GSM transmitter has a unique resource utilization, which dynamically locates the power resource of the carriers based on the traffic and the power needed to transfer. Furthermore, part or all of the carriers are allowed to be shut down to save energy. This feature of the signal adds another dimension of variation and has a significant impact upon DPD design, since not only the average power, but also the probability distribution, PAPR and even bandwidth varies with time. All together with the high requirement on spurious dynamic range (SDR) and adjacent power suppression, which is 60~75 dB, MC-GSM represents a far more unwieldy hurdle for DPD than signals such as WCDMA and LTE, which commonly have a requirement of ACPR of only -55~-45 dB.

The multi-sine signal multiplied to the GSM signal has a strong effect of periodicity, which in turn raises difficulty for delay estimation of MC-GSM signal. Typical delay estimation is to find the peak of the cross-correlation, but the strong periodicity can cause many high pseudo peaks that can interrupt the decision of the true delay. This will require a high sampling frequency and more bits (higher precision) to yield better estimation of cross-correlation function which, however, gives increased hardware complexity and power consumption.

### 2.1.5 Figure of Merit

ACLR and EVM are the most frequently used figure of merit (FoM) for benchmarking the intensity of (nonlinear) distortion. ACLR measures spectral power in the side frequency band, thus it is a direct assessment of the amount of nonlinear distortion. The EVM, on the other hand, is a factor measuring the distortion imposed on symbols and it is seen as a FoM of the in-band distortion of the whole transmitter/receiver. The measurement of both of the

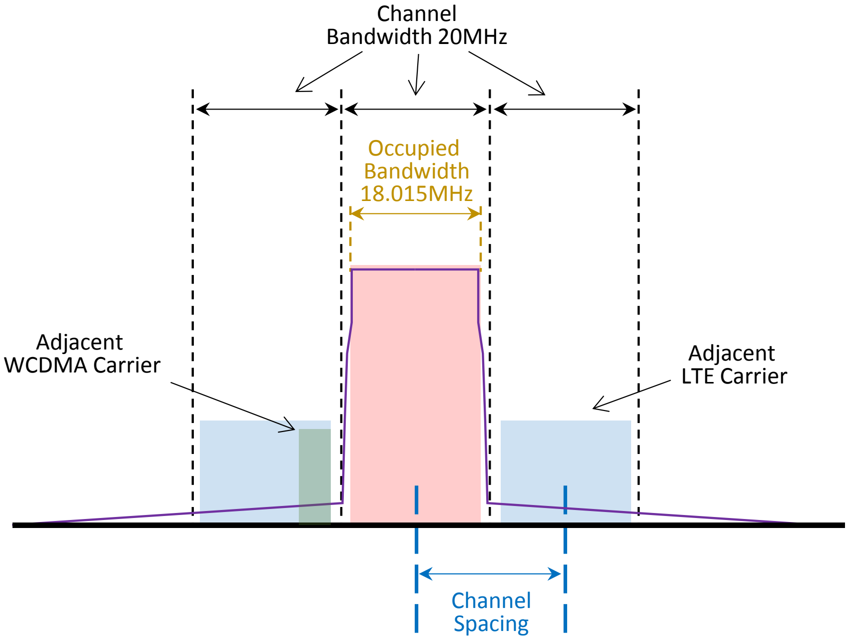


**Fig. 2.5 ACLR measurement for the 20MHz LTE downlink signal.**

two FoMs are regulated by specific standard, and the 3GPP LTE downlink signal is exemplified here to show the measuring procedure.

### *2.1.5.1 ACLR for LTE downlink*

The LTE was developed to backward support predecessor WCDMA system, so that its basic symbol rate inherited the chip rate of WCDMA which is 3.84 Msps (or integer multiples of 3.84). Thus, most of the LTE conformance test for out-of-band emissions are similar in scope and purpose to those for WCDMA. However, LTE spectrum has a rectangular shape that is visually different from WCDMA, which has rounded corners at the band edges. This is due to the different pulse-shaping mechanism and hence the root raised cosine (RRC) filter (with rolloff factor of 0.22) can no longer be applied for LTE to measure the channel power. Actually, no particular type of filter is defined for LTE, thus different filter implementations can be used for LTE transmitter test to optimize either in-channel performance (EVM) or out-of-channel performance (ACLR).

LTE supports six channel bandwidth ranging from 1.4 to 20 MHz. Here, we take the 20MHz downlink LTE as an example to show the measurement of ACLR, which is demonstrated in Fig. 2.5. The signal has a channel bandwidth of 20MHz, which is exclusively used by itself, but the actual occupied bandwidth is 18.015 MHz (15kHz×1201). The ACLR is defined as

$$ACLR(\pm w_{sp}) = \frac{\displaystyle\int_{w=\pm w_{sp}-BW_{adj}/2}^{w=\pm w_{sp}+BW_{adj}/2} \left|X(w)\right|^2 dw}{\displaystyle\int_{w=BW_{ocp}/2}^{w=BW_{ocp}/2} \left|X(w)\right|^2 dw} \tag{2.5}$$

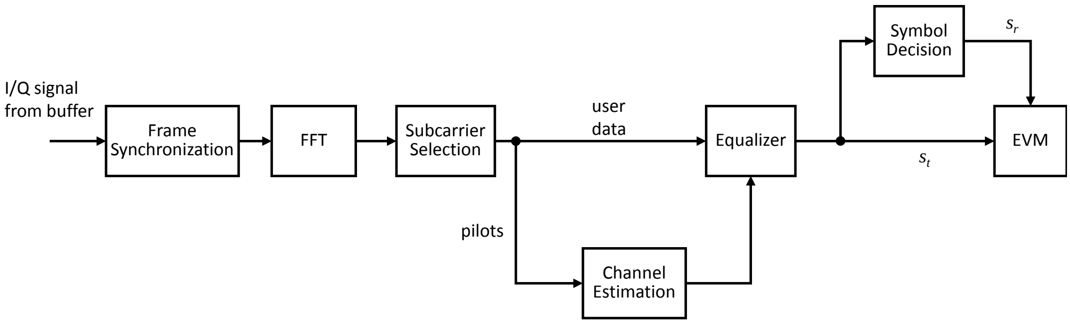


**Fig. 2.6 EVM measurement for the downlink LTE signal.**

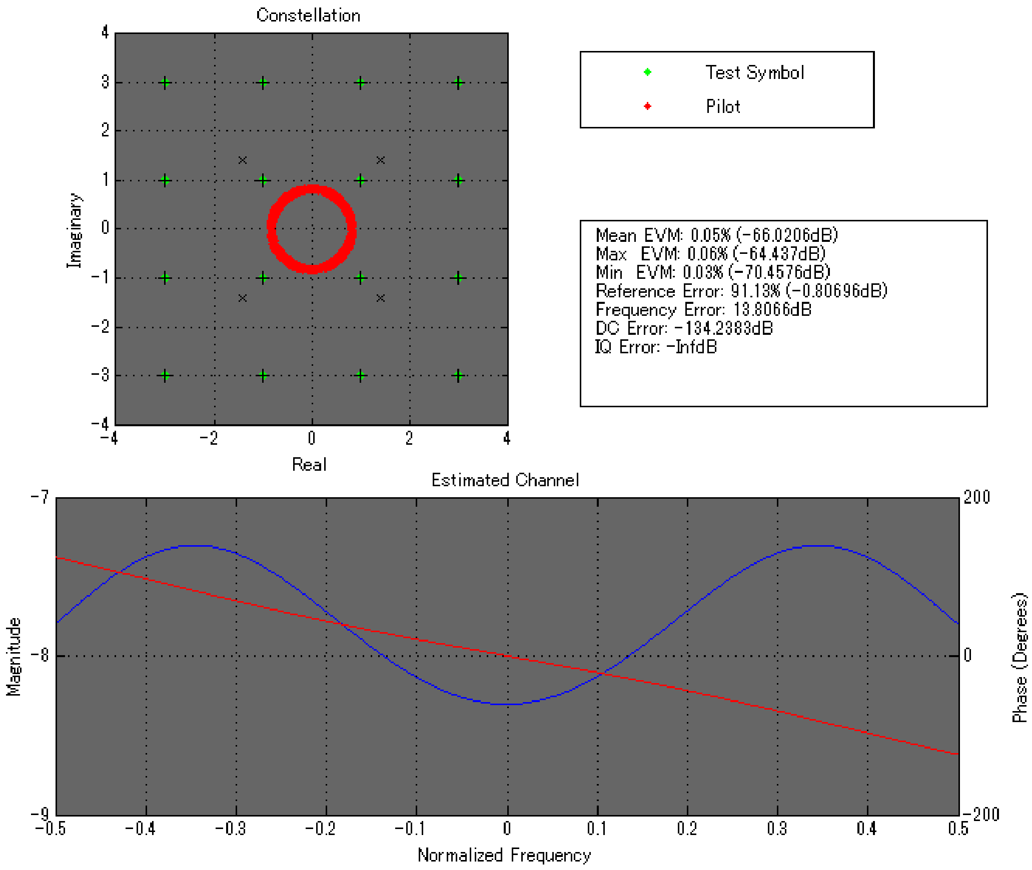
where $BW_{ocp}$ and $BW_{adj}$ denotes the occupied bandwidths for in-channel and adjacent channel, and $W_{sp}$ denotes the channel spacing. $X(w)$ is the Fourier transform of the testing signal, which can be processed by filtering and windowing in order achieve the lowest noise floor. The numerator in (2.5) is the energy of the adjacent channel while denominator represents the energy of the in-band channel. It is worth to note that, the carrier to be transmitted in the adjacent channel may not essentially be the same as the in-channel signal, for instance, if it is a WCDMA signal, the bandwidth for adjacent channel need be modified, as the shaded area in green in Fig. 2.5.

It can be seen that the ACLR is a measurement of the ratio of unwanted spurious and wanted signal, thereby dBc (decibels relative to the carrier) is generally used as its unit.

#### *2.1.5.2 EVM for LTE downlink*

Measuring the EVM is much more complicated as it compasses almost all the procedures needed for the receiver, which include synchronization, equalization, demodulation and etc. This implies that one need the physical layer knowledge of the signal defined by the specifications before making a program to measure the EVM.

A simplified flow chart is depicted in Fig. 2.6 to show the necessary procedures for LTE downlink signal EVM measurement. The EVM is measured on frame basis, so the first step is to find out the position of the P-/S-Sync symbols in time and frequency domain, which



```
Mean EVM: 0.05% (-66.0206dB)
Max  EVM: 0.06% (-64.437dB)
Min  EVM: 0.03% (-70.4576dB)
Reference Error: 91.13% (-0.80696dB)
Frequency Error: 13.8066dB
DC Error: -134.2383dB
IQ Error: -InfdB
```

**Fig. 2.7 User interface of the EVM measurement program produced by the author.**

can also be used to determine the carrier frequency shift. The synchronization may contain several times of rough delay and fine delay to give the best results. The pilots are extracted from the predefined positions of the OFDM symbols, and are used to compensate the linear distortion and scale/rotate the signals to the right place. Decisions are made in the constellation and reference signal $s_r(n)$ is generated. Finally, the EVM is calculated according to the following equation:

$$EVM_{(data)} = \sum_{data} \frac{\|s_r - s_t\|}{\|s_r\|} \tag{2.6}$$

where only the user data is cared.

The author has developed a program for standard-conforming LTE signal generation and EVM test in MATLAB, and its user interface is exemplified in Fig. 2.7. The signal is passed through a frequency-selective channel, which is a Chebyshev filter, as can be inspected from the estimated channel plot at the bottom of the figure. The error caused by I/Q imbalance and carrier LO (local oscillator) leakage can also be estimated by the program, but which is not compensated in EVM test.

### *2.1.5.3 NMSE*

As EVM is difficult to measure, NMSE is frequently used instead. It is defined by the difference between the original signal $x(n)$ and the distorted signal $d(n)$:

$$NMSE(\mathrm{dB}) = 10\log_{10}\left(\sum \frac{|x(\mathrm{n}) - d(\mathrm{n})|^2}{|x(\mathrm{n})|^2}\right) \tag{2.7}$$

where the two signals need to be aligned in time and magnitude/phase. The NMSE is measures the error of the two interpolated signals, thus it contains the distortion of out-of-band contents.

## 2.2 BASICS OF POWER AMPLIFIER

The nonlinear distortion in wireless communication systems mainly originates from the amplifiers including power amplifier (PA) used in transmitter and the low noise amplifier (LNA) used in receiver. Although designs of PA and LNA have different targets, i.e. the PA is to deliver as much power as possible while LNA is to maximize the SDR, the basic non-linear characteristics of the transistors used in the amplifiers have many similarities. Therefore, we only particularize PAs in this section, which is formed with three sub-sections.

### 2.2.1 Power Amplifier Models

#### *2.2.1.1 Equivalent Circuit Modeling*

When designing a circuit, one certainly uses the simulation tool that has the capability to determine circuit performance without making a prototype. This, however, relies on the accuracy of the models, which are parameterized to provide easy access to the characteristics of the circuits. When it comes to the power amplifier, it hence becomes natural to first talk about the equivalent circuit models.

Physical model gives full access to a transistor’s characteristics including almost all the factors such as geometric properties, electromagnetic of gild and substrate, whereas is much too complicated to implement. To cut down the development circle, compact model, which preserves primary features of the physical model and contains less parameters, is frequently used. It can be further categorized, in the scope of different excited signals, into large-signal model and small-signal model. Large-signal model describes transistor’s I-V behavior (DC characteristic) probably from cut-off to breakdown or the thermal behaviors. Small-signal model, on the other hand, gives more precise behavior of the transistor and especially for a certain test signal, which is commonly extracted from the S-parameters. .

The static nonlinearity, in terms of power compression and described by large-signal model, results from output voltages reaching the minimum, or currents approaching the maximum. Starting from the weakly nonlinear operation, the voltage and current swings increase when increasing the output power. The range of currents and voltages, where the model needs to be accurate, therefore increases. However, things change significantly once the voltage dynamically drops below the knee voltage, and the current swings into the saturation region. In this region, the transistor is working in a different mode than in class-A. Therefore, a model that is very accurate in class-A operation can completely fail to predict the compression behavior.

The simulation of the nonlinearity of the PA resorts to both small-signal and large-signal behaviors. For a compact model applied to small-signal and large-signal, the equivalent circuit parameters are subject to transient end voltage. We cannot simply use DC bias as the index, and hope dynamic behavior to be correctly predicted. The small signal model is actually the first term in Taylor series expanded around the DC bias, which defines the slope of the infinite small voltage sweep. True small-signal operation requires only the bias point to be predicted well, together with the S-parameters in the respective frequency range. Of course, this model is restricted to one bias point. If another bias point is set, accuracy cannot be assumed *a priori* without proving first how well the model will match.

### *2.2.1.2 Behavioral Modeling*

The basic strategy taken in digital signal processing for nonlinearity modeling is to emulate the equivalent nonlinearity of the PA (mostly in baseband), instead of predicting the actual I-V relationship and S-parameter. Such special methodology is called behavioral modeling as it only considers the input-output behavior of the circuit. From the point of view of behavioral models, the device under test (DUT) is seen as a black box and the modeling is based kernel fitting, thus the model can provide time-efficient performance evaluation in the system level simulation and is especially suitable for digital predistortion (DPD) and post nonlinear compensation (PNLC).

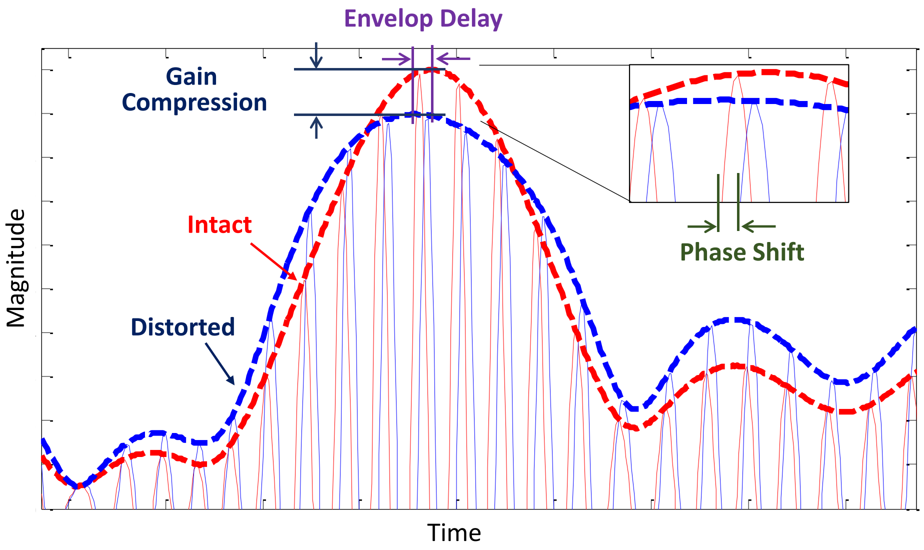
One can observe the compressed magnitude, shifted phase and envelop delay cased by nonlinear PA, as the transient envelop getting large from Fig.2.8. Mathematically, the PA output can be expressed as

$$\tilde{y}(t) = AM\left[A(t)\right]\cos\left(\omega_c t + \theta(t) + PM\left[A(t)\right]\right), \tag{2.8}$$

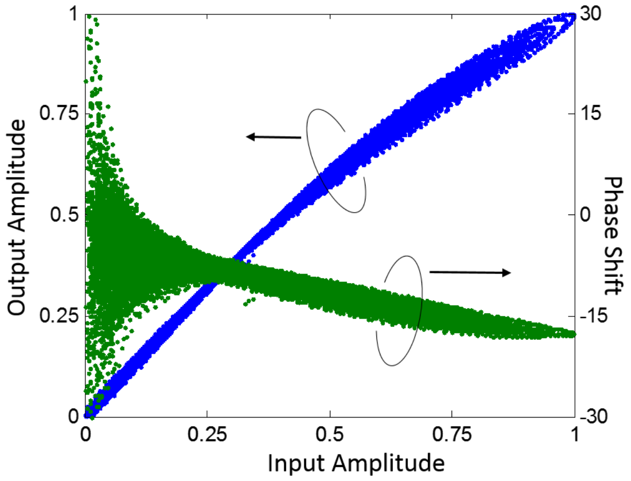
where $AM[\ ]$ and $PM[\ ]$ are nonlinear functions, which indicates that the distortion (as a function of input amplitude) is imposed on the output amplitude and phase shift. The behavioral modeling in baseband is a reflection of the relation of the envelops of the input and output signals, and it can be expressed as

$$y(t) = AM\left[A(t)\right]e^{j\left(\theta(t)+PM\left[A(t)\right]\right)}. \tag{2.9}$$

Behavioral modeling takes into account of the complex envelop irrespective of carriers, hence it yields a model as one manifestation of magnitude-to-magnitude and magnitude-to-



**(a) Demonstration of the time-domain envelops of PA input and output.**



**(b) PA output amplitude and phase shift versus input amplitude plots.**

**Fig. 2.8 Demonstration of the envelops of PA input and output.**

phase, and this is known as AM-AM and AM-PM conversions. Equation (2.9) can be rewritten as a general form of behavioral model in polar form

$$\begin{aligned} y(t) &= A(t)e^{j\theta(t)}\frac{AM\left[A(t)\right]}{A(t)}e^{jPM[A(t)]} \\ &= x(t)\frac{AM\left[A(t)\right]}{A(t)}e^{jPM[A(t)]} \\ &= x(t)G_{AM/PM}\left(A(t)\right) \\ &= x(t)G_{AM/PM}\left(\left|x(t)\right|\right), \end{aligned} \tag{2.10}$$

which is the intact signal multiplied with a nonlinear gain $G_{AM/PM}$.

There are some well-known behavioral models that are widely used in system level simulations, such as Rapp's model and Saleh's model [71]. Such models lack of the ability and accuracy to model various PA's behaviors, as they have only a few controllable parameters and nonlinearly formed parameter spaces. The well-used behavioral models for PA nonlinear modeling and compensation are LUT (look-up table) models, Volterra Series based models and block-oriented models (Wiener/Hammerstein). Review of these models can be found in [72]-[75]. Discussion on different models is not the central topic of this thesis, wherein however, only very simple models such as LUT and memory polynomial are used because the complexity is an important concern here.

It is worth to mention that, behavioral models are commonly extracted by using extra modulated signals $x(t)$ and $y(t)$, and it accordingly only applies when the signal is unchanged. More specifically, changing the signal and especially the average input power and bandwidth can result in dissimilar behaviors. This is because that the behavioral modeling is taken place in limited power and frequency ranges. Therefore, such models can be used in baseband to provide a basic nonlinearity, but lack the extrapolation out of the working range and cannot be used to predict the actual PA characteristics.

In what it follows, it will be seen that the PA's power efficiency is primarily determined by the actual voltage and current waveform, and the conduction angle of the PA also has an effect on the degree of nonlinearity.

### 2.2.2 Power Amplifier Operation Modes and Power Efficiency

The transistor can be seen as a voltage control current source (VCCS) regardless of parasitics, if FET (field effect transistor) is assumed. The DC I-V characteristics of the transistor is described by the large signal model, where for instance, the Ids-Vds plots with different Vgs are shown in Fig. 2.9. This plot is commonly used to determine the bias point or operation mode of the amplifier. If the bias point is set to operate the amplifier in class-A, the signals on the DC bias can swing along the load line with a 360 degree of conduct angle, while for class-B operation mode, only half of the signal can be conducted.

The operation mode determines the efficiency of the PA that can be achieved as well as the amount of harmonic distortion. The power efficiency can be calculated by

$$\eta = \frac{P_{output}}{P_{supply}} = \frac{P_{supply} - P_{consume}}{P_{supply}} \tag{2.11}$$

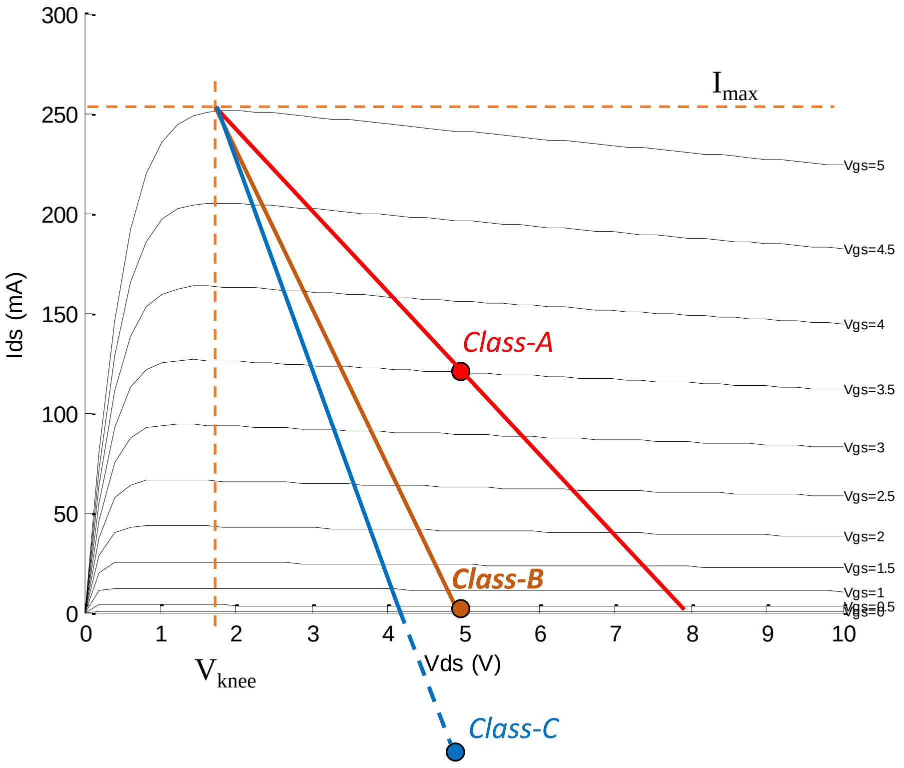


**Fig. 2.9 Demonstration of the I-V plot for Materka-Kacprzak model.**

which is very straightforward that the efficiency is determined by the power supply and the power consumed by the transistor. Thus, reducing the consuming power is a natural result to enhance the efficiency of the PA. The power consumed by the transistor is determined by the drain-source current and voltage. Thus, we can expect the efficiency by examining the waveform of the current and voltage. As the digitally modulated signal has a slow varying envelop compared to the carrier (see equation (2.3)), the signal applied to the PA can be seen as a sine wave in a short time scale.

The I/V waveforms for class-A are sine signals only with 180 degree difference in phase, as shown in Fig. 2.10. The waveforms overlap with each other through the whole period, so that the power consumed by the transistor takes up half of the power supply. With the operation in class-B, the overlap of the current and voltage is halved, and power consumption can be significantly cut down. The class-C reduces the power consumption by further shrinking the overlaps of current and voltage. By applying pulse-width modulated envelop or properly utilizing the harmonic components, one can obtain a rectangular voltage and a half-wave current, which have no overlapping, as shown in right bottom of Fig. 2.10. The former operation is referred as class-D and the latter is class-F, and both of which can theoretically achieve 100% efficiency as the power consumption caused by the transistor is subtly avoided. Especially, the class-F, which is achieved by controlling (opening or shorting) the harmonic

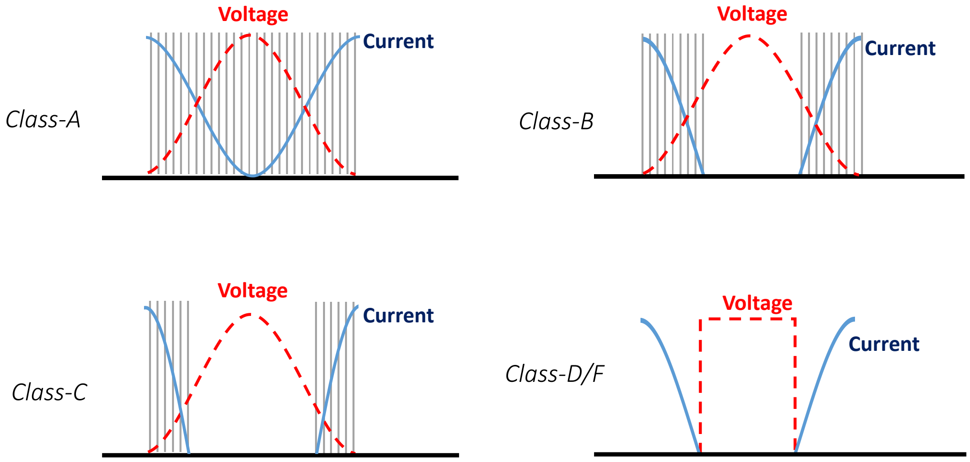


**Fig. 2.10 Demonstration of the current and voltage waveform for class-A, class-B, class-C and class-D/F amplifiers.**

impedances [76][77], has formed the basic ideas for designing a high-efficiency PA in recent years.

## 2.2.3 Power Amplifier Nonlinearity

The power amplifier nonlinearity cause distortions that can be observed by the occurrence of harmonic and the distorted signal envelop. As the harmonic components can be easily removed by filtering and only the envelop carries useful information for communication (see equation (2.3)), the distortion imposed on envelop deserves more attentions. In this section static nonlinearity including AM-AM and AM-PM conversion, as well as dynamic nonlinearity known as memory effect are presented.

### *2.2.3.1 AM-AM Distortion*

To understand the basic nonlinear mechanism of the AM-AM conversion, we first assume ideal cut off and operation of the transistor here. Resorting to Fig. 2.10, one can see the output waveform and the spectrum for PAs in different operation modes. For class-A, as the output current is totally conducted, no distortion can be observed from the spectrum in right hand. The class-B has a half wave output and shows a spectrum free from odd-order harmonics, resulting in a clean signal in desired band. Therefore, operating close in class-B is crucial to design a linear PA. The class-AB, notwithstanding larger conduction angle than class-B, shows harmonics of both even and odd orders thereby the desired signal is contaminated by odd-order distortions. Class-C PA has conduction angle less than 180 degrees, thus there are very likely a portion of envelops being clipped, see left bottom of Fig. 2.11.

The input envelop amplitude versus output envelop amplitude is defined as the AM-AM characteristic of the PA, which for class-A, class-AB, class-B and class-C are plotted in Fig. 2.12. The ideal cut off shows no impact on the class-A and class-B PAs, but gives arise of

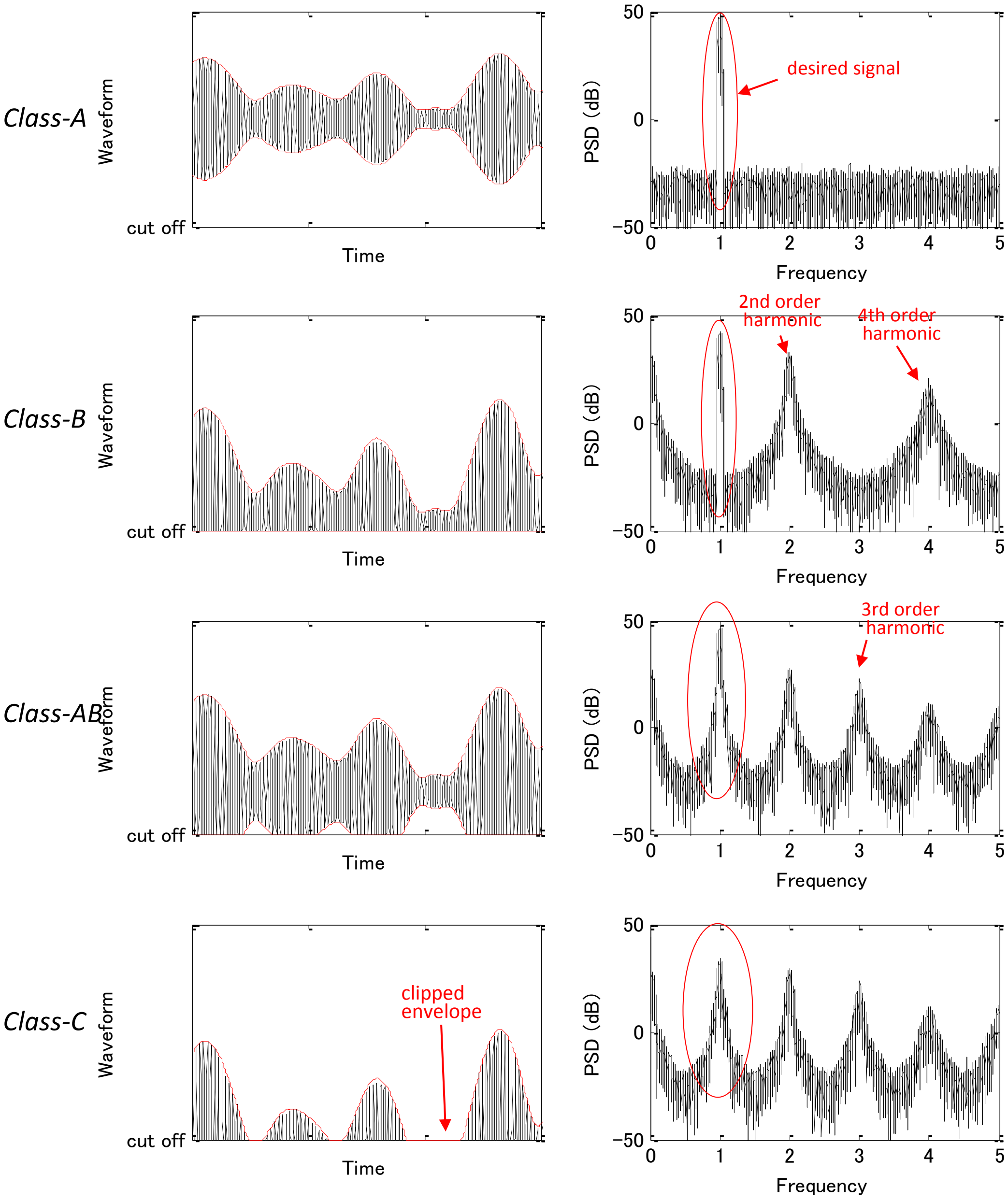


**Fig. 2.11 Demonstration of the time domain waveform and spectrum for class-A, class-B, class-AB and class-C amplifier.**

expanded gain to class-AB and hard clipping to class-C. The change in gain yields inevitable nonlinear distortion, as can also been seen from the spectral plots in Fig. 2.11, and this implies that the class-AB and class-C PAs are essentially nonlinear. However, the cut off effect for class-C results in a baseband signal with hard clipped amplitude, which is unable to recover. Therefore, nonlinear compensation for class-C mode is seldom touched and hardly found its applications for linear amplifying.

As signal envelop getting strong, it leads to larger swing along the load line shown in Fig.

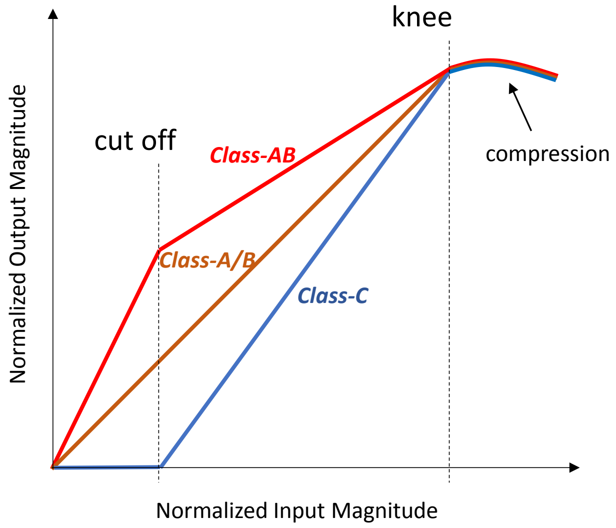


**Fig. 2.12 Demonstration of the AM-AM plots for class-A, class-AB, class-B and class-C amplifiers.**

2.9, the signal can be distorted by a so-called *knee* effect. When the signal cross the bound set by $V_{knee}$, the output current drops, being responsible for the output power compression, as shown in Fig. 2.12. Therefore, even in class-A or class-B mode, the PA also shows nonlinearity when input signal is strong.

To sum up, the AM-AM distortion is ascribed to two factors: the operation mode that sets the amount of waveform to be cut off; and the knee effect that causes gain compression when input is strong. For real transistors, the cut off effect will yield a more smoothed AM-AM curve for class-AB, making it easier for nonlinear compensation. Moreover, as already mentioned before, as the behavioral model has no enclosure of signal carrier, the operation modes and power efficiency, which are defined by the voltage and current waveforms, cannot be speculated by behavioral modeling.

### *2.2.3.2 AM-PM Distortion*

AM-PM conversion is a distortion existing in PAs that causes input amplitude dependent phase shift of the output [78]. The interpretation of AM-PM, such as rotated intermodulation distortion (IMD) [79][80], quasi-memory effect [81], has been widely known. It was also verified in [82] that the AM-PM is originated from the nonlinear capacitances that generate a quadrature current that causes the nonlinear phase shift. It is also noticed that the AM-AM conversion modulates the Miller capacitance caused by feedback capacitance, which pulls the phase response of the input matching network if the impedance transformation ratio is high.

Despite of what has really caused the AM-PM, the result for that is the dissimilar delays imposed on the different harmonic components. Here, for example, we assume the nonlinearity of the PA follows such a nonlinear function with 3rd order distortion:

$$\tilde{y}(t) = a_1 A(t)\cos\left(\omega_c t + \theta(\mathrm{t})\right) + a_3 A^3(t)\cos^3\left(\omega_c(t-\tau) + \theta(\mathrm{t})\right), \tag{2.12}$$

where $a_1$ and $a_3$ are the gains for elemental and 3rd order harmonic, respectively, and $\tau$ is a constant representing the delay imposed on 3rd order harmonic. After rewriting the products and powers of the trigonometric functions it becomes

$$\begin{aligned} \tilde{y}(t) &= a_1 A(t)\cos\left(\omega_c t + \theta(\mathrm{t})\right) + \frac{3}{4} a_3 A^3(t)\cos\left(\omega_c(t-\tau) + \theta(\mathrm{t})\right) \\ &+ \frac{1}{4} a_3 A^3(t)\cos\left(3\omega_c(t-\tau) + 3\theta(\mathrm{t})\right). \end{aligned} \tag{2.13}$$

Obviously, only the component at $\omega_c$ is cared, thereby

$$\tilde{y}(t,\omega_c) = a_1 A(t)\cos\left(\omega_c t + \theta(\mathrm{t})\right) + \frac{3}{4} a_3 A^3(t)\cos\left(\omega_c(t-\tau) + \theta(\mathrm{t})\right), \tag{2.14}$$

whose corresponding baseband signal can be written as

$$y(t) = a_1 A(t)\,\mathrm{e}^{j\theta(\mathrm{t})} + \frac{3}{4} a_3 A^3(t)\,\mathrm{e}^{j(\theta(\mathrm{t}) - \omega_c\tau)}. \tag{2.15}$$

Recalling (2.1), it arrives at

$$y(t) = a_1 x(t) + \frac{3}{4} a_3\,\mathrm{e}^{-j\omega_c\tau} \left|x(t)\right|^2 x(t), \tag{2.16}$$

which has a complex coefficient for the 3rd order operator. The resulting phase shift is the angle defined by

$$PM\left(\left|x(t)\right|\right) = \angle\left(\frac{y(t)}{x(t)}\right) = \angle\left(a_1 + \frac{3}{4} a_3\,\mathrm{e}^{-j\omega_c\tau}\left|x(t)\right|^2\right), \tag{2.17}$$

which alternatively

$$PM\left(\left|x(t)\right|\right) = \angle\left(a_1 + \frac{3}{4} a_3 \left|x(t)\right|^2 \left[\cos(\omega_c\tau) - j\sin(\omega_c\tau)\right]\right), \tag{2.18}$$

thus

$$PM\left(\left|x(t)\right|\right) = -\arctan\frac{\frac{3}{4} a_3 \sin(\omega_c\tau)\left|x(t)\right|^2}{a_1 + \frac{3}{4} a_3 \cos(\omega_c\tau)\left|x(t)\right|^2}. \tag{2.19}$$

It can be seen from the above equation that two parameters have determined the AM-PM conversion, which are $a_3$ and $\tau$. Generally, $\tau$ is a temporal parameter determined by memory elements and $a_3$ is a large-signal parameter defined by the nonlinear admittance of the transistor. Moreover, $a_3$ is dependent on the bias voltage $V_{GS}$, thus the PA in different operation modes can also yield dissimilar AM-PM conversion, as shown in Fig.2.13.

The cut off effect for different operation mode also has an effect on the AM-PM and it is

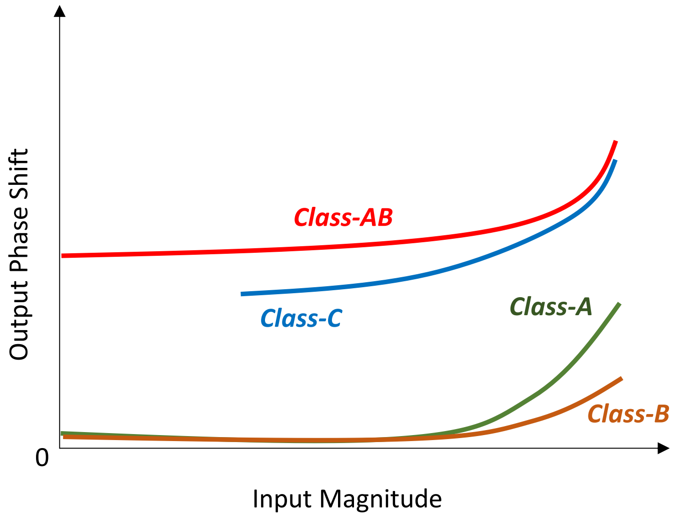


**Fig. 2.13 Demonstration of the AM-PM plots for class-A, class-AB, class-B and class-C amplifiers.**

generally transferred to a constant phase shift, as depicted in Fig. 2.13. The class-A and class-B have nearly zero constant phase shift, but class-A suffers more AM-PM distortion as it has larger portion of envelop enter the nonlinear region for achieving the same average output power as class-B. The class-C has ambiguous phase that is intractable in the cut off region and also shows a constant phase shift as the class-AB.

#### *2.2.3.3 Memory Effect*

The memory effect is categorized into thermal effect and electrical effect, but in recent years, the term *memory effect* has almost become the synonym of electrical effect. It is because of the slow variation of thermal effect that, comparing to the variation speed of the signal envelop, the electrical effect is more evident and thermal effect becomes negligible when the signal bandwidth surpasses 100 kHz [83].

The electrical memory effect (which is called *memory effect* hereinafter) originates from the frequency-dependent impedance in and is the results of transistor parasitic, matching network and other large memory components including capacitances and inductances. However, it cannot be simply interpreted as the unflatness in frequency domain (linear distortion), because of the mixing and remixing of harmonic components. The recognition that memory effects dominates the PA nonlinear behavior to a large extent has been a major roadblock for the nonlinear compensation. With the progress has been made to transistors so far, the memory effect can be minimized through semiconductor process and circuit development.

The direct results of memory effect are the dispersion of AM-AM/AM-PM conversion and unequal out-of-band power emissions for upper and lower bands. One can observe the memory effect from the AM-AM plot abstained from a two-tone test in the left of Fig.2.14, which is similar to a hysteresis. More specifically, there have two respective gains when the instantaneous signal amplitude is rising and dropping. For a signal with an occupied bandwidth (with infinite frequency components) rather than two frequency components, the AM-AM plot becomes rather dispersive as can be seen in the right of Fig.2.14.

From the point view of behavioral modeling, memory effect can be interpreted as a dynamic nonlinearity thereby it can be modeled as using the analysis methods under the framework of Volterra series. Discussion on the memory effect is out of the scope of this thesis and a detailed introduction can be found in [84].

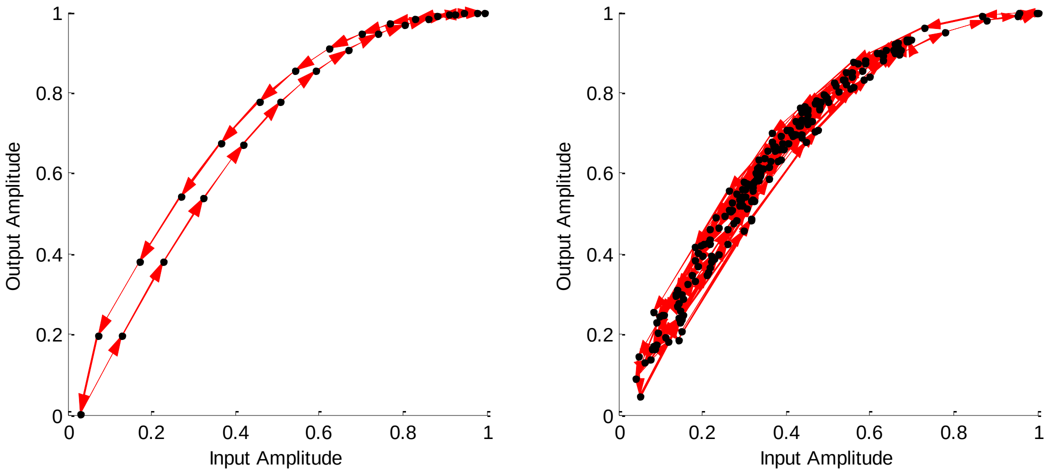


**Fig. 2.14 Demonstration of the AM-AM plot in the presence of memory effect, where the signal is a two-tone signal (left) and digitally modulated signal with occupied bandwidth (right).**

# Chapter 3.

# Mathematical Formulations of Nonlinear Modeling and Theories of Inverses

In this chapter, the general mathematical formulations on nonlinear modeling and the theories of inverse methodologies using polynomial models are presented. Section 3.1 presents the polynomial model, where it is given an algebraic architecture thereby it can be studied using general matrix theory. The involved issues include the model's linear regressive structure, orthogonalized basis space as well as the method for subspace selection (model pruning).

Either digital predistortion (DPD) or post nonlinear compensation (PNLC) involves the procedure to construct a relationship representing the inverse of the target. Thus, it becomes natural to consider the theories for inverting a given nonlinear system. In section 3.2, a general recursive method is introduced to construct the inverse for a nonlinear function (nonlinear system). A method for inverting a given polynomial, which is optimum in a sense of minimized mean square error (MSE), is followed. In section 3.3, the mathematical formulations set up for coefficients extraction for DPD and PNLC are also conceptually introduced.

## 3.1 POLYNOMIALS FOR NONLINEAR MODELING

### 3.1.1 Representing the Nonlinearity in Baseband

Since the nonlinearity caused by power amplifier (PA) and other circuits is very weak, so that they can be modeled with Volterra series, which has been widely applied for modeling a weak nonlinear dynamic system:

$$y(n)=\sum_{p=1}^{P}\sum_{i_1=0}^{M}\cdots\sum_{i_p=0}^{M}h_p(i_1,\cdots,i_p)\prod_{q=1}^{p}x(\mathrm{n}-i_q), \tag{3.1}$$

where $h_p(i_1,\cdots,i_p)$ is called the $p$th-order Volterra kernel. Here $P$ determines the maximum order of nonlinearity and $M$ stands for the maximum memory depth. It is indubitably much too complicated and most of the operators can be discarded to form a more compact model.

A numerical method to prune the model will be introduced in 3.14, and here we will first consider a simple case, which is a $Q$th-order power series:

$$\begin{aligned}\tilde{y}(t) &= b_1\tilde{x}(t) + b_2\tilde{x}^2(t) + \cdots + b_Q\tilde{x}^Q(t) \\ &= \sum_{q=1}^{P} b_q\tilde{x}^q(t).\end{aligned} \tag{3.2}$$

Here, we use tilde to denote the signal in passband, as the one defined in (2.3).

Before proceeding with the discussion, it is important to clarify the relation and difference between two cases: the passband polynomial that describes complete RF signal, and low-pass equivalent polynomial that process only the envelop information. Output of a nonlinear passband system can be expressed by (3.2), provided that the system obeys the finite-order polynomial model without memory. It can be expressed by its complex envelop

$$\tilde{x}(t) = \Re\{x(t)e^{2\pi f_c t}\} = \frac{1}{2}(x(t)e^{2\pi f_c t} + x^*(t)e^{-2\pi f_c t}), \tag{3.3}$$

where $f_c$ is the carrier frequency. By replacing $\tilde{x}(t)$ in (3.2) with the expression of (3.3), the complex envelop of the real output signal is given by

$$y(t) = \sum_{i=0}^{(P+1)/2} \frac{b_{2i+1}}{2^{2i}} \binom{2i+1}{i+1} |x(t)|^{2i} x(t) = \sum_{i=0}^{(P+1)/2} c_i |x(t)|^{2i} x(t). \tag{3.4}$$

Note that the baseband expression considers only odd order terms since harmonics of even-order do not introduce distortions in the frequency band of interest. Nevertheless, as shown in [85], introducing even-order terms in (3.4) can be beneficial to power amplifier modeling and predistortion. Thus it can be further written as, if even-order terms are introduced:

$$y(t) = \sum_{i=1}^{P} c_i |x(t)|^{i-1} x(t). \tag{3.5}$$

It is worth to emphasize that the even order terms are different from the terms in (3.2). They are introduced deliberately and have no relation with the passband model.

With the relation developed before, we can also get the baseband expression for Volterra series:

$$y(n) = \sum_{p=1,3,5\cdots}^{P} \sum_{i_1=0}^{M} \cdots \sum_{i_p=0}^{M} \left[ h(i_p i_p) x(n-i_1) \prod_{q=1}^{(p-1)/2} x(n-i_{2q}) x^*(n-i_{2q+1}) \right]. \tag{3.6}$$

A well-established PA model is the memory polynomial, which as a pruned version of Volterra series has less number of coefficients and an easy-implementing structure:

$$y(n) = \sum_{q=1}^{M} \sum_{p=1}^{N} a_{pq} |x(n-q+1)|^{p-1} x(n-q+1). \tag{3.7}$$

This model gives good performance in practical application, hence it will be used throughout this thesis.

### 3.1.2 Algebraic Architecture of the Polynomial Model: Linear Regression

The polynomial models shown above, in spite of the nonlinear operators it contains, has a linear regressive architecture, which thus can be written as a generalized form of:

$$y(n)=\sum_{i=1}^{N}a_i k_i\left(x(n)\right). \tag{3.8}$$

where $a_i$ is the coefficient for the $i$th nonlinear operator $k_i(\ )$. It is instructive to clarify here that although $k_i(\ )$ is nonlinear function, the model has a linear architecture as $k_i(\ )$ is linear respective to the coefficient $w_i$. Therefore, $k_i(\ )$ is called linear regression and (3.8) is thus a linear regressive model.

By stacking the regressions together in a matrix given by

$$K\left(x(n)\right)=\left[k_1\left(x(n)\right)\quad k_2\left(x(n)\right)\quad\cdots\quad k_N\left(x(n)\right)\right], \tag{3.9}$$

which forms the basis space for the model, the output can be written as

$$y(n)=K\left(x(n)\right)\mathbf{A}, \tag{3.10}$$

where

$$\mathbf{A}=\left[a_1\quad a_2\quad\cdots\quad a_N\right]^T \tag{3.11}$$

is the coefficient vector. If the signal $x(n)$ has a finite length, say $L$, the model can be expressed in a matrix form:

$$\begin{aligned}&\mathbf{x}=\left[y_1\quad y_2\quad\cdots\quad y_L\right]^T,\\&\mathbf{y}=\left[y_1\quad y_2\quad\cdots\quad y_L\right]^T,\\&\mathbf{X}=K(\mathbf{x})=\left[k_1(\mathbf{x})\quad k_2(\mathbf{x})\quad\cdots\quad k_N(\mathbf{x})\right],\\&\mathbf{y}=\mathbf{XA}.\end{aligned} \tag{3.12}$$

The linear regressive architecture of the polynomial model allows the use of linear estimator for coefficient calculation. One well-known linear estimator is the least squares (LS), where for instance if **d** is the desired output we can use

$$\mathbf{A}=\left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{d} \tag{3.13}$$

to estimate the parameter.

Therefore, the procedure of modeling utilizing polynomial models is actually a linear optimization problem, except for some special cases encountered in DPD, as will be discussed in 3.4.1.

### 3.1.3 Representing the Model with Orthogonal Basis

**Table 3-1 Pseudo code for generating the projection matrix**

```
Initialization :  $\boldsymbol{\Psi}=\mathbf{X}$    $\mathbf{U}=\begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix}_{(L\times L)}$
Update :  Loop1: k=2:L
              $\mathbf{G}=\begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix}_{(L\times L)}$
              Loop2: m=k:L
                  $\mathbf{G}(k,m)=-\frac{\boldsymbol{\varphi}_k^H\boldsymbol{\varphi}_m}{\boldsymbol{\varphi}_k^H\boldsymbol{\varphi}_k}$
                  $\boldsymbol{\varphi}_m=\boldsymbol{\varphi}_m+\mathbf{G}(k,m)\boldsymbol{\varphi}_k$
              $\mathbf{U}=\mathbf{UG}$
              $\boldsymbol{\varphi}_k=\frac{\boldsymbol{\varphi}_k}{\|\boldsymbol{\varphi}_k\|}$    $\mathbf{U}(k,:)=\frac{\mathbf{U}(k,:)}{\|\boldsymbol{\varphi}_k\|}$
          $\boldsymbol{\varphi}_L=\frac{\boldsymbol{\varphi}_L}{\|\boldsymbol{\varphi}_L\|}$    $\mathbf{U}(L,:)=\frac{\mathbf{U}(L,:)}{\|\boldsymbol{\varphi}_L\|}$
Adjust :  $\mathbf{U}=\frac{\mathbf{U}}{u_{11}}$    $\boldsymbol{\Psi}=\frac{\boldsymbol{\Psi}}{u_{11}}$
```

If the signal is wide sense stationary, and its distribution is known *a priori*, we can find a matrix to orthogonalize the polynomial. Define the covariance matrix

$$\mathbf{R}=E[\mathbf{X}^H\mathbf{X}] \tag{3.14}$$

where the superscript $^H$ denotes complex conjugate transpose and $E[\ ]$ is the expectation value. It can be transformed to a unit matrix **I**, if a projection matrix is introduced,

$$E[(\mathbf{XU})^H\mathbf{XU}]=E[\boldsymbol{\Psi}^H\boldsymbol{\Psi}]=\mathbf{I}. \tag{3.15}$$

The projection matrix **U** recasts the basis space **X** in a space that all the subspaces are orthogonal to each other. Here, we use

$$\boldsymbol{\Psi}=\mathbf{XU}=\begin{bmatrix}\boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_N\end{bmatrix} \tag{3.16}$$

to denote the orthogonalized space, where

$$\begin{aligned}
\boldsymbol{\varphi}_1 &= u_{11}\mathbf{x}_1 \\
\boldsymbol{\varphi}_2 &= u_{12}\mathbf{x}_1 + u_{22}\mathbf{x}_2 \\
&\vdots \\
\boldsymbol{\varphi}_N &= u_{1N}\mathbf{x}_1 + u_{2N}\mathbf{x}_2 + \cdots + u_{NN}\mathbf{x}_N
\end{aligned} \tag{3.17}$$

is the operators of the orthogonal model and

$$\mathbf{U} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1N} \\ 0 & u_{22} & \cdots & u_{2N} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & u_{NN} \end{bmatrix} \tag{3.18}$$

is a upper triangular matrix.

One can enjoy a lot of benefits from orthogonal models, such as better numerical stability and faster convergence rate. For example, the author in [86] derived the projection matrix **U** in closed form for uniform distributed signal, which however, is derived based on memoryless polynomial and lacks the generality of applicate to other models. In fact, as the polynomial is a linear regressive model, the basis space can be easily orthogonalized by well-established mathematical algorithms such as Householder/Givens rotation [87], Gram-Schmidt orthogonalization [88], etc. Here, we present a method to derive **U** based on modified Gram-Schmidt procedure [89], which can be numerically superior to the classic Gram-Schmidt orthogonalization, and its pseudo code is summarized in Table 3.1.

### 3.1.4 Model Pruning using Subspace Selection

Much effort has been undertaken to prune the Volterra series, yielding a complexity-reduced model. This is to remove the terms with less contribution while preserving the dominate terms. The memory polynomial, for instance, is the result taken from the diagonal direction of the Volterra parameter space [90]. In practical applications, as presented in Section 1.1.3.3, only simple model can be afforded and even the memory polynomial still be too complicated. Therefore, model pruning is crucial to reduce the complexity, as a result of less computational load for model coefficients extraction.

Model pruning is intrinsically a subspace selection problem, i.e. selecting dominant columns from the whole basis space **X**, and several mathematical methods have been developed to accomplish this task, which are based on either singular value decomposition (SVD) or QR factorization [91].

Now we first return to the result shown in the last subsection where we have developed the orthogonalized basis space:

$$\boldsymbol{\Psi} = \mathbf{XU}. \tag{3.19}$$

If we use the above orthogonal space (orthogonal model) to perform the estimation with LS

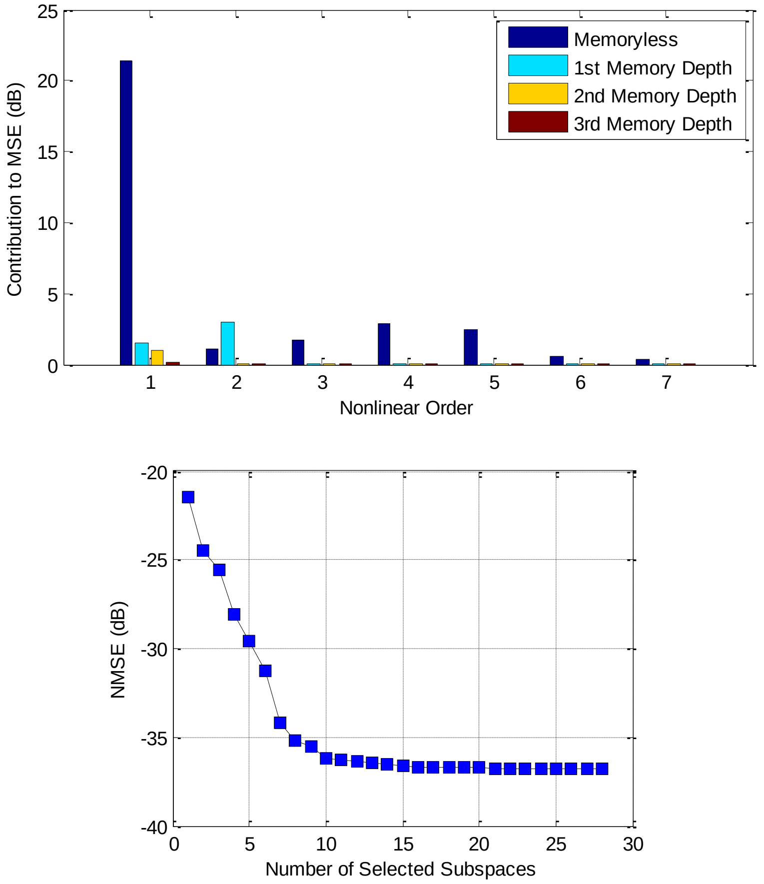


**Fig. 3.1 Subspace selection results for a memory polynomial.**

defined in (3.13), we will have:

$$\bar{\mathbf{W}} = \left(\mathbf{\Psi}^H \mathbf{\Psi}\right)^{-1} \mathbf{\Psi}^H \mathbf{d}. \tag{3.20}$$

Recalling (3.15), there yields:

$$\bar{\mathbf{W}} = \mathbf{\Psi}^H \mathbf{d}, \tag{3.21}$$

which reveals a merit of using orthogonal model that the matrix inverse can be removed. The residual error of LS, namely mean square error (MSE), can be written as

$$\begin{aligned}
\zeta &= \left\| \mathbf{\Psi}\bar{\mathbf{W}} - \mathbf{d} \right\|^2 \\
&= \left(\mathbf{\Psi}\bar{\mathbf{W}} - \mathbf{d}\right)^H \left(\mathbf{\Psi}\bar{\mathbf{W}} - \mathbf{d}\right) \\
&= \bar{\mathbf{W}}^H \mathbf{\Psi}^H \mathbf{\Psi}\bar{\mathbf{W}} - \bar{\mathbf{W}}^H \mathbf{\Psi}^H \mathbf{d} - \mathbf{d}^H \mathbf{\Psi}\bar{\mathbf{W}} + \mathbf{d}^H \mathbf{d} \\
&= \bar{\mathbf{W}}^H \bar{\mathbf{W}} - \bar{\mathbf{W}}^H \mathbf{\Psi}^H \mathbf{d} - \mathbf{d}^H \mathbf{\Psi}\bar{\mathbf{W}} + \mathbf{d}^H \mathbf{d}.
\end{aligned} \tag{3.22}$$

Substituting (3.21) into the above equation it yields:

$$\begin{aligned}\zeta &= \left(\mathbf{\Psi}^H\mathbf{d}\right)^H \mathbf{\Psi}^H\mathbf{d} - \left(\mathbf{\Psi}^H\mathbf{d}\right)^H \mathbf{\Psi}^H\mathbf{d} - \mathbf{d}^H\mathbf{\Psi}\mathbf{\Psi}^H\mathbf{d} + \mathbf{d}^H\mathbf{d} \\ &= \mathbf{d}^H\mathbf{d} - \mathbf{d}^H\mathbf{\Psi}\mathbf{\Psi}^H\mathbf{d} \\ &= \mathbf{d}^H\mathbf{d} - \sum_{i=1}^{N}\left\|\boldsymbol{\varphi}_i^H\mathbf{d}\right\|^2.\end{aligned} \tag{3.23}$$

Therefore, the $\boldsymbol{\varphi}_i$ $(i=1,2,\ldots,N)$ that result in largest $\left\|\boldsymbol{\varphi}_i^H\mathbf{d}\right\|^2$ are the dominant terms with the largest contribution to MSE.

Our task is to select the dominate subspaces from the basis space **X** instead of **Ψ**, thus there needs some extra efforts to refactor the **X** from the reduced **Ψ**. This is a matrix factorization updating problem [91] and is not discussed furthermore in this thesis.

To demonstrate the effectiveness of subspace selection for model pruning, we used a memory polynomial defined in (3.7) to model an actual Doherty amplifier, where *N* and *M* were set to 7 and 4, respectively, giving a total basis number of 28. One can see the contribution of each basis of the model from up of Fig. 3.1, which provide an intuitionistic chart for pruning the model. One can also see from the bottom of the figure that leaving only less than 10 basis can obtain comparative performance as using the full model, indicating that most of basis in the memory polynomial are reluctant and have trivial contribution to MSE.

## 3.2 RECURSIVE INVERSE METHOD

### 3.2.1 Recursive Inverse Structure

Here, we consider a continuous function f(x) that meets such conditions: monotonous, say

$$f(x_1) > f(x_2) \qquad (x_1 > x_2) \tag{3.24}$$

or

$$f(x_2) > f(x_1) \qquad (x_2 > x_1) \tag{3.25}$$

and zero-crossing at 0:

$$f(0) = 0. \tag{3.26}$$

The first condition set above is obvious because only monotonous function (at least in the region of interests it is monotonous) is invertible. The second condition ensures that the function can be expanded about the 0 point and it contains no DC component.

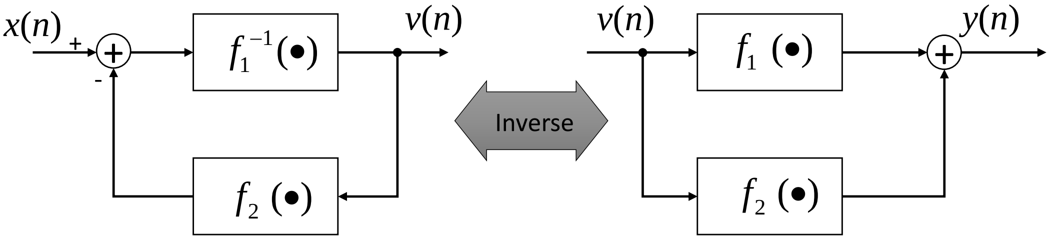


**Fig. 3.2 Demonstration of the inverse of a nonlinear function.**

For any linear regressive model, we can express it as

$$y = f(x) = f_1(x) + f_2(x), \tag{3.27}$$

the inverse function which can be expressed as:

$$f^{-1}(x) = f_1^{-1}\left(f(x) - f_2\left(f(x)\right)\right), \tag{3.28}$$

as has been shown in [20] already. The above function gives a recursive structure to invert the given function. Expressing $f(x)$ with addition of two functions of arbitrary form, so it is beneficial to select $f_1(x)$ that is easy to find the inverse function. Equation (3.28) can be easily proved by taking inspection of Fig. 3.2, where

$$v(n) = f_1^{-1}\left(x(n) - f_2\left(v(n)\right)\right) \tag{3.29}$$

$$y(n) = f_1\left(v(n)\right) + f_2\left(v(n)\right), \tag{3.30}$$

and it yields

$$y(n) = x(n) - f_2\left(v(n)\right) + f_2\left(v(n)\right) = x(n). \tag{3.31}$$

This shows that with the feedback (recursive) structure shown in left hand in Fig. 3.2, perfect inverse can be constructed since it finally produces the output equaling to the input. Note that the recursive structure produces infinite orders of nonlinearity, indicating that even for a nonlinearity with finite order, its perfect inverse is a nonlinearity with infinite order.

## 3.2.2 Inverse using Iterative Feedback

Fig. 3.2 provides an easy-to-implement structure to construct the exact inverse of a given nonlinear function by portioning the target function into $f_1(x)$ and $f_2(x)$, and placing $f_2(x)$ in the feedback path and inverse of $f_1(x)$ in the forward path. In practical applications, we can choose an easily invertible $f_1(x)$, such as $x$.

Note that in digital domain, at least one register representing one sample of delay must be inserted in the feedback path, but if $f(x)$ is memoryless function, the recursive structure becomes unrealizable. To overcome this problem, the author has proposed an iterative feed

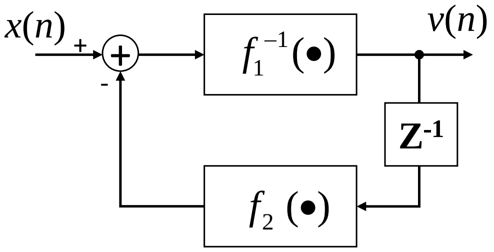


**Fig. 3.3 Realizable recursive structure**

back method to approximate the ideal recursive structure [20]. Fig. 3.3 shows a structure that can be realized in practice, which compare to the one in left hand in Fig. 3.2 contains one delay tap in feedback path, thereby the output becomes:

$$v(n)=f_1^{-1}\left(x(n)-f_2\left(v(n-1)\right)\right), \tag{3.32}$$

which approximates (3.29) if the delay is infinitely small. Thus, (3.32) can give good inverse performance if the sampling rate is very high (corresponding to small unit sampling period). However, increasing the sampling rate may not always be feasible, and accordingly a method is developed to synthesize the feedback loop even with low sampling rate. This is accomplished by iterations:

$$\begin{aligned} v^{(k)}(n) &= f_1^{-1}\left(x(n)-f_2\left(v^{(k-1)}(n)\right)\right), \\ v^{(0)}(n) &= x(n) \end{aligned} \tag{3.33}$$

which is the $k$-th iteration result and the initial value $v^{(0)}(n)$ can be set to $x(n)$ for simplicity. Note that the second term in the right hand is iteratively replaced by the output in the last iteration. This method can be easily extended to memory nonlinear function inverses, e.g. the memory polynomial, as shown in [20].

## 3.2.3 Expanding Iterative Feedback

The method shown above is difficult to be implemented in high-speed because of the need for iteration and its recursive structure as well as the need for iteration. Here, a method [92] for obtaining the inverse for a given continuous and monotonous nonlinear function is proposed. The method can be easily implemented as it has an elegant structure for the inverse construction. More specifically, the proposed inverse can be easily constructed, if the function of the target nonlinear system is known. Furthermore, the inverse can be fully pipelined, allowing high-speed real time implementation.

Now we define a function, which is given as

$$g(x)=f_1^{-1}\left(x-f_2(x)\right) \tag{3.34}$$

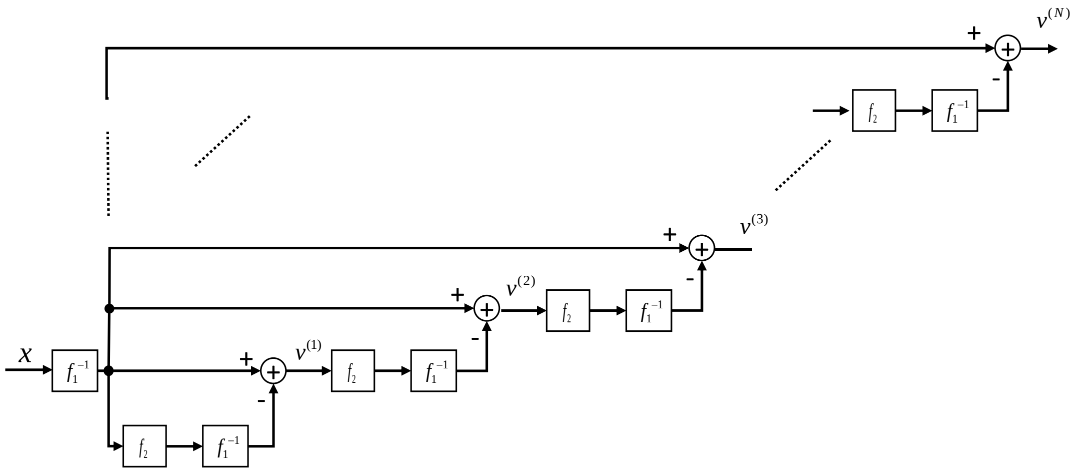


Fig. 3.4 **Expanded iterative feedback structure.**

then the inverse function given in (3.11) can be approximated by connecting a number of *g*( ) in tandem, say

$$v^{(N)} = \underbrace{g\left(\cdots g\left(g\left(g(x)\right)\right)\right)}_{N} \tag{3.35}$$

where the *g*( ) is repeated *N* times. This function is plotted in Fig. 3.4, from which it can be seen that the diagram contains no feedback path so that it is possible to be implemented fully pipelined. If *N* is large enough (3.35) can infinitely approach (3.11), which is the actual exact of *f*( ).

## 3.3 POLYNOMIAL INVERSES

There are several techniques for inverting a given polynomial. Schetzen [93] proposed "*p*th-order inverse" method to invert a Volterra series which is known as the most general polynomial for representing weak nonlinear systems with memory. In order to increase the speed and decrease the complexity for constructing the Volterra kernels, Sarti et al. proposed the recursive technology for the *p*th-order inverse [94]. The limitation of the *p*th-order inverse is that it can only remove the nonlinear distortion introduced by the terms up to *p*th-order and leaves residual higher order distortion. It results in ineffective inverses in some circumstances. Tsimbinos et al. [95] derived the method to invert the power series based on Chebyshev polynomial and Hermite polynomial, which provides superior performance to that of *p*th-order inverse.

A new algorithm is proposed that gives the optimum baseband polynomial inverse with a limited order [96][97]. The algorithm employs orthogonal basis that is predetermined from the distribution of input signal. The major difference of the proposed method from the conventional methods is that it focuses on minimizing the mean square error, while conventional works aims to remove the higher distortions of *p*th-order or above, not to minimize the whole distortion power.

This algorithm is based on the assumption that the nonlinear system can be represented by a baseband polynomial. When a stationary input signal and its probability distribution are given, the algorithm can find the optimum coefficients of the baseband inverse polynomial with a limited order. With orthogonal basis, all orders of distortion components can be simultaneously suppressed. The optimum inverse coefficients are obtained by directly manipulating the coefficients of the given polynomial. Another great advantage of the approach is that the accuracy of suppressing higher order distortion can be controlled by a truncation factor. This results in high flexibility in terms of practical implementation and enables compromise between its performance and complexity.

## 3.3.1 *p*th-order Inverse and its Limitation

The *p*th-order inverse method proposed in [93] is under the general frame of Volterra series. Its basic idea is that compensation of Volterra series using another Volterra series of *p*th-order removes the first *p*th-order composite kernels.

Consider a polynomial *f* that is followed by another polynomial *h*. If *h* is the inverse of *f*, the whole system's response is linear.

Let *f*(*x*) is a 3rd order polynomial that is already known as

$$y = f(x) = a_1x + a_2x^2 + a_3x^3 \tag{3.36}$$

If a 2nd-order polynomial which is given by

$$z = h(y) = b_1y + b_2y^2 \tag{3.37}$$

is employed as the inverse, the output of the two polynomials in tandem with respect to the input *x* is

$$\begin{aligned} z &= b_1(a_1x + a_2x^2 + a_3x^3) + b_2(a_1x + a_2x^2 + a_3x^3)^2 \\ &= b_1a_1x + (b_1a_2 + b_2a_1^2)x^2 + (b_1a_3 + 2b_2a_1a_2)x^3 + (b_2a_2^2 + 2b_2a_1a_3)x^4 \\ &\quad + 2b_2a_2a_3x^5 + b_2a_3^2x^6 \end{aligned} \tag{3.38}$$

which is a 6th-order polynomial. According to the *p*th-order inverse, the coefficients $b_1$ and $b_2$ of the inverse polynomial in (3.37) can be obtained by solving for the following equations

$$\begin{cases} b_1a_1 = G_{lin} \\ b_1a_2 + b_2a_1^2 = 0 \end{cases} \tag{3.39}$$

where $G_{lin}$ is the targeted linear gain.

Note that the *p*th-order inverse can only remove the 2nd-order monomial in (3.38) while higher order monomials are generated. Therefore *p*th-order inverse is, in this sense, not the optimum inverse of the given polynomial. This example also reveals a fact that a polynomial of finite order cannot completely remove finite order distortions, and residual distortions of higher orders always remain.

### 3.3.2 Polynomial Inverse

In order to derive the optimum polynomial inverse, three steps are necessary. The first is to write the explicit expression for the composite polynomial, and then transform it to corresponding orthogonal polynomial. The inverse can be obtained that minimizes the mean square error of the output signal.

#### *3.3.2.1 Composition of two polynomials in tandem*

Composition means to expand the expression of two polynomials in tandem. Given two tandem polynomials such as

$$y = f(x) = \sum_{i=1}^{N} a_i x^i \tag{3.40}$$

$$z = g(y) = \sum_{j=1}^{M} b_i y^j \tag{3.41}$$

Substituting (3.40) in (3.41) results in

$$z = \sum_{j=1}^{M} b_j \left(\sum_{i=1}^{N} a_i x^i\right)^j \tag{3.42}$$

Composition of two polynomials involves polynomial multiplications. For instance, the $m$th-order term in (3.42) is

$$\left(\sum_{i=1}^{N} a_i x^i\right)^m \tag{3.43}$$

The expanded coefficients of (8) can be calculated through vector convolutions

$$\mathbf{A}^{(m)} = \underbrace{\mathbf{A} * \mathbf{A} * \cdots * \mathbf{A}}_{m \quad terms} \tag{3.44}$$

where $\mathbf{A}=[a_1\ a_2\ \ldots\ a_N]^T$ is the coefficient vector. Then (3.44) can be expressed as $\mathbf{X}_1\mathbf{A}^{(m)}$, where $\mathbf{X}_1=[x^m, x^{m+1}, \ldots, x^{mN}]$.

We write the polynomial of (3.40) in the form of $y$=$\mathbf{XA}$, where $\mathbf{X}=[x, x^2, \ldots, x^N]$. In the similar way, Eq. (3.41) can also be expressed as $z$=$\mathbf{YB}$, where $\mathbf{Y}=[y, y^2, \ldots, y^M]$, and $\mathbf{B}=[b_1\ b_2\ \ldots\ b_M]^T$. We stack the coefficients for (3.43) in a vector and zeros are paddled for alignment

$$\mathbf{p}_m = [\underbrace{0, 0, \ldots, 0}_{m-1 \quad zeros}, (\mathbf{A}^{(m)})^T]^T \tag{3.45}$$

which is a column vector with length of $mN$+$m$-2. Arraying them into a matrix it gives

$$\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_3, \cdots, \mathbf{p}_M] \tag{3.46}$$

which is a ($MN$+$M$-2)×$M$ matrix. Then the composite polynomial can be expressed as

$$z = \mathbf{X}_p \mathbf{P}_{(MN\times M)} \mathbf{B} \tag{3.47}$$

where $\mathbf{X}_p=[x, x^2, \ldots, x^{MN}]$, and **P** is truncated to $MN\times M$, because there are only $MN$ terms in the composite polynomial.

### *3.3.2.2 Polynomial Inverse*

As shown in Fig. 2, the two systems connected in tandem are linear if $z=\mathbf{G}_{lin}x$. Compensating a polynomial $f$ in (3.40) with another polynomial $h$ in (3.41) means to linearize the composite polynomial in (3.47),

$$\mathbf{PB}=\mathbf{G} \tag{3.48}$$

where $\mathbf{G}=[\mathbf{G}_{lin}, 0, 0\ldots, 0]^T$ is a column vector with $MN$ elements. Calculating the inverse polynomial coefficients requires solving the **B** in (3.48). Note that there are $M$ elements in **B**. Therefore **P** needs to be truncated for conformability.

If the coefficient vector **B** is solved as follows,.

$$\mathbf{B}=(\mathbf{P}_{(M\times M)})^{-1}\mathbf{G}_{(M\times 1)} \tag{3.49}$$

where **P** is truncated to $M\times M$ so that the first $M$th order distortions are removed, then, Eq. (3.49) is equivalent to $p$th-order inverse method.

To find the coefficient **B**, we define the cost function as the mean square error

$$\begin{aligned}\xi &= E[|G_{lin}x-z|^2]=E[|\mathbf{XG}-\mathbf{XPB}|^2] \\ &= E[\mathbf{B}^H\mathbf{P}^H\mathbf{X}^H\mathbf{XPB}+\mathbf{G}^H\mathbf{X}^H\mathbf{XG}-\mathbf{G}^H\mathbf{X}^H\mathbf{XPB}-\mathbf{B}^H\mathbf{P}^H\mathbf{X}^H\mathbf{XG}]\end{aligned} \tag{3.50}$$

By applying the orthogonal transform to the composite polynomial, (3.50) yields

$$\begin{aligned}\xi = E[&\mathbf{B}^H\mathbf{P}_{orth}{}^H(\mathbf{XU})^H\mathbf{XUP}_{orth}\mathbf{B}+\mathbf{G}^H(\mathbf{XU})^H\mathbf{XUG} \\ &-\mathbf{B}^H\mathbf{P}_{orth}{}^H(\mathbf{XU})^H\mathbf{XUG}-\mathbf{G}^H(\mathbf{XU})^H\mathbf{XUP}_{orth}\mathbf{B}]\end{aligned} \tag{3.51}$$

According to (3.15),

$$\begin{aligned}\xi &= (\mathbf{B}^H\mathbf{P}_{orth}{}^H\mathbf{P}_{orth}\mathbf{B}+\mathbf{G}^H\mathbf{G}-\mathbf{B}^H\mathbf{P}_{orth}{}^H\mathbf{G}-\mathbf{G}^H\mathbf{P}_{orth}\mathbf{B}) \\ &= |\mathbf{P}_{orth}\mathbf{B}-\mathbf{G}|^2=|\mathbf{U}^{-1}\mathbf{PB}-\mathbf{G}|^2\end{aligned} \tag{3.52}$$

is signal-independent. The least square solution for minimizing $\xi$ is

$$B=\left(\mathbf{P}_{orth}^H\mathbf{P}_{orth}\right)^{-1}\mathbf{P}_{orth}^H\mathbf{G}_{(MN\times 1)} \tag{3.53}$$

where

$$\mathbf{P}_{orth}=(\mathbf{U}_{(MN\times MN)})^{-1}\mathbf{P}_{(MN\times M)} \tag{3.54}$$

In (3.53) all the distortion components are suppressed simultaneously because the composite coefficients matrix **P** contains all the $MN\times M$ terms.

Next, define a truncation factor $K$ to further decrease the computation complexity. It yields

$$\mathbf{B} = \left(\Omega^H \Omega\right)^{-1} \Omega^H G_{(K\times 1)} \tag{3.55}$$

where

$$\Omega = (U_{(K\times K)})^{-1} P_{(K\times M)} \tag{3.56}$$

The truncation factor $K$ denotes the maximum order of distortion that can be suppressed. The range of $K$ is from $M$ to $MN$. Unlike the $p$th-order inverse that can only remove the first $p$th-order terms, the algorithm proposed in (3.55) and (3.56) can suppress higher order terms by adjusting $K$. Performing the proposed algorithm requires the transform matrix **U**, hence the distribution of the signal should be known *a priori*. If this algorithm is applied in a real time circumstance, e.g. a digital predistortion, an offline procedure can be done to generate **U** before running the system.

### 3.3.2.3 Baseband Polynomial Inverse

If a baseband polynomial with only odd-order terms is considered, its inverse will be quite easy following the algorithm introduced in the last section, because there is a direct pattern to map the passband polynomial to the baseband counterpart. However, inverse of a baseband polynomial with even-order terms requires modification of the algorithm.

Suppose that two baseband polynomials are given by

$$y = f(x) = \sum_{i=1}^{N} a_i |x|^{i-1} x \tag{3.57}$$

$$z = g(y) = \sum_{i=1}^{M} b_i |y|^{i-1} y \tag{3.58}$$

Same as the passband counterpart, the first step is composition of the two polynomials

$$z = \sum_{j=1}^{M} b_j \left| \sum_{i=1}^{N} a_i |x|^{i-1} x \right|^{j-1} \sum_{i=1}^{N} a_i |x|^{i-1} x \tag{3.59}$$

The $m$th-order term in (3.59) is

$$\left| \sum_{i=1}^{N} a_i |x|^{i-1} x \right|^{m-1} \sum_{i=1}^{N} a_i |x|^{i-1} x \tag{3.60}$$

if $m$ is odd, it can be written as

$$\underbrace{\left(\sum_{i=1}^{N} a_i |x|^{i-1} x\right)^{*} \left(\sum_{i=1}^{N} a_i |x|^{i-1} x\right) \cdots}_{m-1 \quad terms} \sum_{i=1}^{N} a_i |x|^{i-1} x \tag{3.61}$$

herein the coefficients of which can be calculated by defining

$$\mathbf{A}^{(M)} = \underbrace{\mathbf{A} * \mathbf{A}^{*} * \cdots * \mathbf{A} * \mathbf{A}^{*}}_{m-1 \quad terms} * \mathbf{A} \tag{3.62}$$

The problem left is how to cope with the even order terms, as which generate infinite order of distortions and they cannot be expanded easily. When *m* is even, (3.61) yields

$$\left|\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right| \underbrace{\left(\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right)^{*} \left(\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right) \cdots}_{M-2 \quad terms} \sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x \tag{3.63}$$

which always leaves an absolute term

$$\left|\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right| = \sqrt{\left(\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right)^{*} \left(\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right)} \tag{3.64}$$

Absolution of the complex is, in fact, a severe nonlinear operation because of its discontinuity. If only the absolute term is considered which is expressed as:

$$\begin{aligned}
&\left|\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right| \\
&= \sqrt{\left(a_1 x + a_2 |x| x + a_3 |x|^2 x + \cdots + a_N |x|^{N-1} x\right)^{*} \left(a_1 x + a_2 |x| x + a_3 |x|^2 x + \cdots + a_N |x|^{N-1} x\right)} \\
&= |a_1||x| \sqrt{1 + \frac{(a_1^{*} a_2 + a_1 a_2^{*})|x|}{|a_1|^2} + \frac{(a_1^{*} a_3 + a_1 a_3^{*} + |a_2|^2)|x|^2}{|a_1|^2} + \cdots}
\end{aligned} \tag{3.65}$$

Considered for a weak nonlinear system, $|a_i|$, $i > 1$, are always much smaller than $|a_1|$, and $x$ is assumed to be small, thus it can be approximated as:

$$\begin{aligned}
\left|\sum_{i=1}^{N} a_i \,|\,x\,|^{i-1}\, x\right| &\fallingdotseq |a_1||x| + \frac{1}{2}\left(\frac{(a_1^{*} a_2 + a_1 a_2^{*})}{|a_1|}|x|^2 + \frac{(a_1^{*} a_3 + a_1 a_3^{*} + |a_2|^2)}{|a_1|}|x|^3 + \cdots\right) \\
&\fallingdotseq |a_1||x| + \frac{1}{2}\left(\frac{(a_1^{*} a_2 + a_1 a_2^{*})}{|a_1|}|x|^2 + \frac{(a_1^{*} a_3 + a_1 a_3^{*})}{|a_1|}|x|^3 + \cdots\right)
\end{aligned} \tag{3.66}$$

Consequently, the coefficients of the absolution term can be expressed as

$$\mathbf{A}_{abs} = \frac{\mathbf{A}^{*} a_1 + \mathbf{A} a_1^{*}}{2|a_1|} \tag{3.67}$$

Therefore, the coefficients of (3.60) are

$$\mathbf{A}^{(M)} = \begin{cases} \underbrace{\mathbf{A} * \mathbf{A}^{*} * \cdots \mathbf{A} * \mathbf{A}^{*}}_{M-1 \quad terms} * \mathbf{A} & M \text{ is odd} \\ \underbrace{\mathbf{A} * \mathbf{A}^{*} * \cdots \mathbf{A} * \mathbf{A}^{*}}_{M-2 \quad terms} * \mathbf{A}_{abs} * \mathbf{A} & M \text{ is even} \end{cases} \tag{3.68}$$

and (3.60) can be expressed in a manner of matrices: $\mathbf{X}''\mathbf{A}^{(M)}$, where $\mathbf{X}''=[x, |x|x, |x|^2x, \ldots, |x|^{M-1}x]$.

Note that the even-order and odd-order terms are handled separately, because the even order terms cannot be expanded as the odd-order terms. Then all the works left are the same as passband polynomial inverses in last subsection.

## 3.4 METHODOLOGIES FOR DIGITAL PREDISTORTION AND POST NONLINEAR COMPENSATION

After discussing general nonlinear modeling and inverses so far, now we will return to the issues concerning digital predistortion (DPD) and post nonlinear compensation (PNLC) in due course. From the point of view of system level modeling, the DPD is to find the pre-inverse of a target nonlinear block while PNLC is to find the post-inverse, so that when they are set upstream/downstream of the target nonlinear block, the response as a whole becomes linear. In this section, the discussions about DPD and PNLC mostly regards to model coefficients extraction issues are presented without touching on detailed methods, to portray a conceptual framework for these two techniques.

### 3.4.1 Coefficients Extraction for Digital Predistortion

The predistorter (PD) is the pre-inverse of the PA, and for instance if use g( ) representing the function of PD and f( ) the function of PA, the final output is the result of the two functions in tandem:

$$y = f(v) = f\left(g(x)\right). \tag{3.69}$$

We can see that

$$g(x) = f^{-1}\left(y\right), \tag{3.70}$$

as simple as it sound, we need the inverse function $f^{-1}(\,)$, for which however, $f(\,)$ is unknown to us. Therefore, it becomes a natural consequence to first estimate the PA model $f_m(\,)$, as diagrammatically shown in Fig. 3.5, in the light of the following criteria:

$$\min E\left(\left|y - y_{\mathrm{mod}}\right|^2\right) = \min E\left(\left|y - f_m\left(v\right)\right|^2\right), \tag{3.71}$$

and then obtain the inverse of $f_m(\,)$. Obviously, it contains two steps: PA modeling and inverse estimation. As linear regressive model is used, there is no much difficulty for coefficients extraction. In addition, the PA model and inverse are not necessarily to be taken place consecutively, which means that the signals can be buffered and coefficients extraction can be done with a slower rate rather than the signal transmission rate, hence allowing flexible DPD implementation.

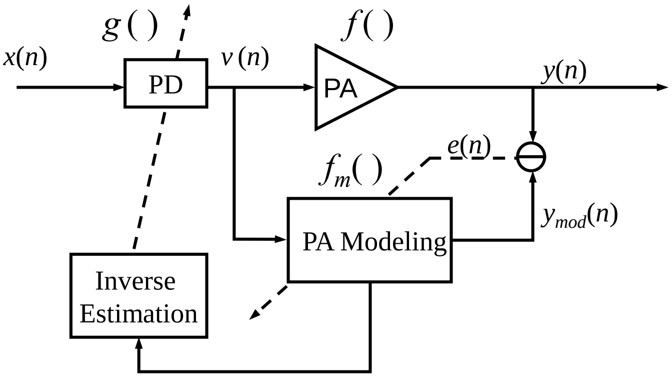


Fig. 3.5 DPD based on modeling and inverse.

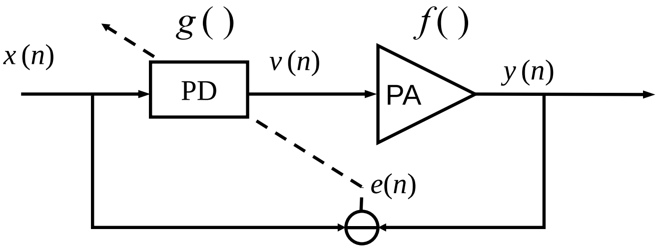


Fig. 3.6 DPD based on direct learning.

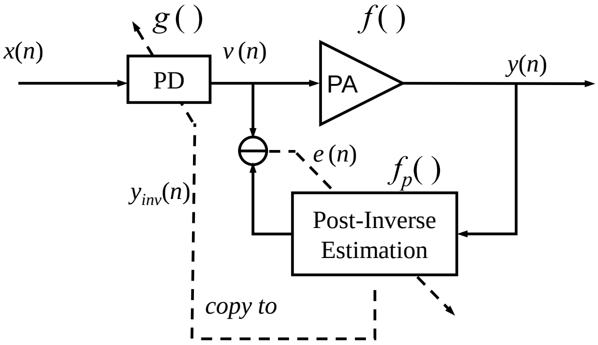


Fig. 3.7 DPD based on indirect learning.

The PD coefficients can be estimated directly without modeling the PA, see Fig. 3.6, which is to apply such a criteria

$$\min E\left(|x-y|^2\right)=\min E\left(\left|x-f\left(g(x)\right)\right|^2\right). \tag{3.72}$$

Note that even if $g(\ )$ is a linear regressive model, (3.72) represents a nonlinear optimization problem because $g(\ )$ is in tandem with a nonlinear function of $f(\ )$. Thus, one can use nonlinear optimization algorithms like the one used in Chapter 5.3 to fulfil the task in due, and it is also shown in Chapter 4.2.2 that we can also apply linear estimator to this problem even though *direct learning* itself is a nonlinear process. Nonlinear optimization may not be convex thereby is difficult to find the global optimum solution. In practice, the robustness of *direct learning* is also tempt to be affected by many external factors, thus, albeit architectural simplicity, it is more difficult to implement.

*Indirect learning*, see Fig. 3.7, which estimates the post-inverse instead and copy the extracted coefficients to the PD, is an alternative way for direct inverse estimation. The procedure of optimization towards to

$$\min E\left(\left|v-f_p\left(f(v)\right)\right|^2\right)=\min E\left(\left|v-f_p\left(y\right)\right|^2\right) \tag{3.73}$$

is a linear process. It has better robustness, as well as high flexibilities of implementation. Thus it is quite popular in DPD community and can be found utilizations in most of the academic papers.

### 3.4.2 Coefficients Extraction for Post Nonlinear Compensation

Unlike the DPD, the intact signal is unknown to receiver, thus the PNLC is to estimate the post-inverse with only distorted signal. As the error cannot be defined as the difference of intact signal and distorted signal as the one in DPD, PNLC represents a blind nonlinear modeling problem.

Observing that nonlinearity causes spectral regrowth around the band of interest, we can define the out-of-band spectral signal as error. Recalling Bussgang's theorem [98] given as

$$y=\lambda x+err, \tag{3.74}$$

where $x$ is the original signal, *err* is the distortion and $\lambda$ is a constant. The spectrum of *err*, falls onto $x$, causing both in-band and out-of-band distortions, as shown in left of Fig. 3.8. The above equation is heuristic as it reveals a fact that the original signal and distortion are independent and separable. Therefore, when both the distorted signal $y$ and intact signal $x$ are available, the error is to subtract $x$ from $y$, yielding the error as shown in Fig. 3.8.

For blind nonlinear modeling, the out-of-band spectral signal is normally obtained with a notching filter, which is equivalently a high-pass filter (HPF) in baseband, as demonstrated in Fig. 3.9. Compared with Fig. 3.8, the error obtained by HPF is defective, especially lacks of the in-band information. Consider that the memory effect is generally seen as the unflatness of the frequency response, blind modeling is unable to handle memory nonlinearity. Fortunately, the received signal does not require a high SDR as the transmitter as it is hampered by noise channel, thus using a memoryless model for PNLC is sufficient to most applications.

Finally, it is worth to mention that, PNLC utilizing blind nonlinear modeling is also a linear optimization problem as the filtering (using FIR) is a linear operation, giving us much confidence for designing an optimal compensator.

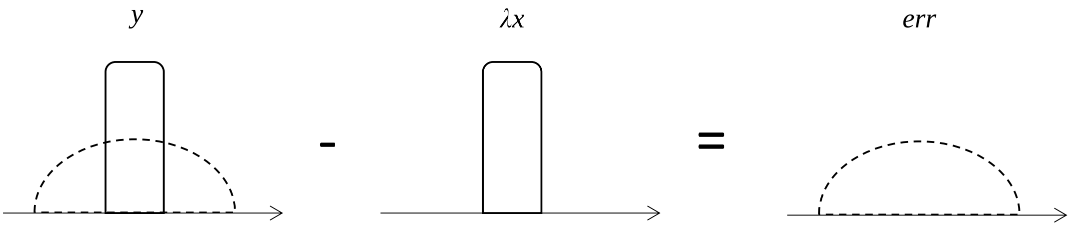


**Fig. 3.8 Demonstration of the error in frequency domain, when the intact signal is available.**

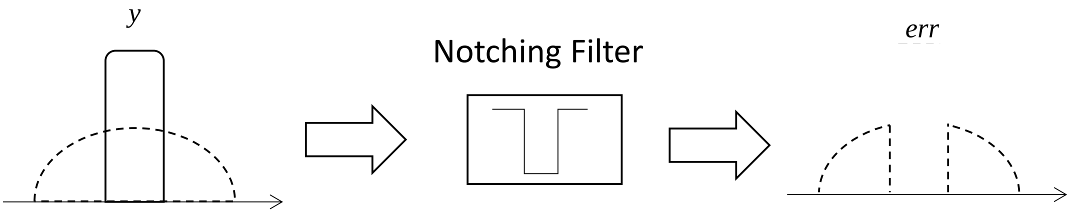


**Fig. 3.9 Demonstration of the error in frequency domain, when the intact signal is not available.**

# Chapter 4.

# Instrumentation and Implementation

The experiments taken place for digital predistortion (DPD) and post nonlinear compensation (PNLC) validation include instrumental measurement [99] and FPGA (field programmable gate array) implementation. To perform the experiment, a transmitter circuitry that includes basic digital-to-analog conversion and up-conversion is needed to transfer the signal from baseband to radio-frequency (RF), and a receiver circuitry, which includes down-conversion and analog-to-digital-conversion, serving as the reversed functional block of the transmitter. Passed through data conversion and RF links, some post-processing are needed to align the signals, which includes time delay alignment, I/Q impairment compensation, linear equalization and etc. This chapter focuses on the issues that must to be resolved in practical instrumental measurement and FPGA implementation.

## 4.1 INSTRUMENTAL MEASUREMENT

### 4.1.1 Experiments for DPD and PNLC Validation

The DPD is the inverse of PA obtained through comparison of the intact signal and distorted signal in baseband, as have seen in Section 3.4.1. The intact baseband signal is the signal to transmit and is already known to the DPD, but the distorted signal needs an extra receiver (RX) path to transfer the PA output to baseband. The DPD output is converted to RF through a transmitter (TX) path, and is directly fed into the PA. The signal bandwidth and PAPR are two most important factors affecting the behavior of a PA. Hence, a measurement platform with wide bandwidth and high accuracy is necessary to fully access the characteristic of a PA. The TX and RX are phase-locked to maintain the phase relationship between the input signal and output signal. To accurately observe the characteristic, or behavior, of the PA, the RX is required to have enough bandwidth and dynamic range. The bandwidth of the signal acquisition path, in general, should be at least five times of the signal bandwidth to accurately capture the behavior of the PA due to the spectral regrowth caused by the intermodulation products. When the experimental setup is used for predistortion, on top of the bandwidth requirement, both signal generation and acquisition paths should have a dynamic range that provides better adjacent chancel leakage ration (ACLR) than the required linearity. Moreover, the peripheral circuits contained in the TX and RX should have high reliability to ensure that no extra distortions are introduced in the measurement.

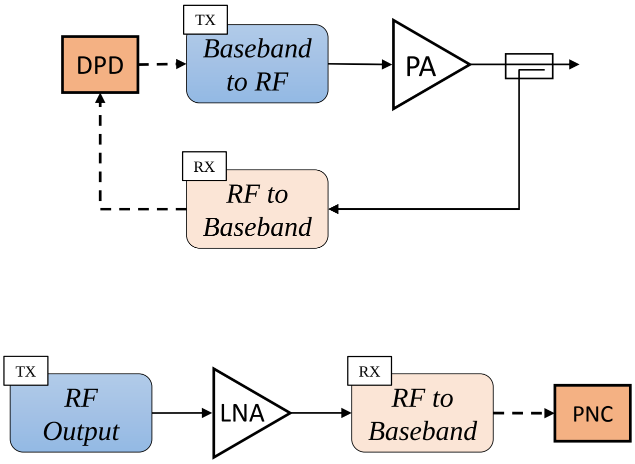


**Fig. 4.1 Outline of test bed for DPD and PNLC.**

The PNLC, on the other hand, is to manipulate the distorted and the intact signal is not required. Therefore, the test bed for PNLC is simpler than DPD, e.g. the TX and RX need not to be phase-locked, and the delay and magnitude mismatching do not affect the operation of PNLC.

The outline of the test bed for DPD and PNLC is illustrated in Fig. 4.1, where the only difference is the device under test (DUT) between the transmitter and receiver: the PA for DPD and LNA for PNLC. Thus, it is of no surprise that the observation data measured from the DUT should have high accuracy and reliability, because different signal sets used to abstract the model may lead to dissimilar parameters.

### 4.1.2 SG-SA Test Bed

The test bed shown in Fig. 4.2 is formed by commercial instruments, which are SG (signal generator) and SA (signal analyzer), and a computer with necessary communication interfaces and signal processing software. The SG can provide a streaming signal, which however is repeated with an extra period depending on the maximum memory of the SG. The SG and SA can share a sine signal, normally from 10 MHz to 100 MHz, as a common

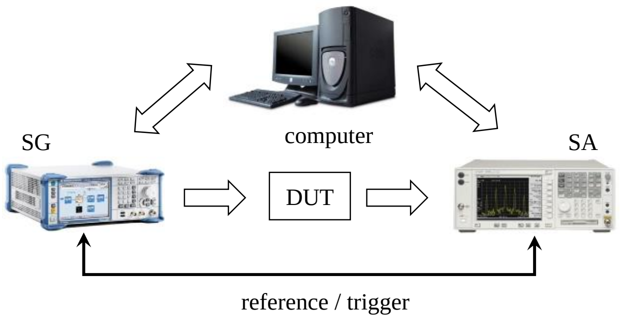


**Fig. 4.2 Depiction of the test bed formed by commercial instruments.**

**Table 4-1 The list of used instruments**

| Instruments | Description |
|---|---|
| N5182/N5183 | Vector signal generator with maximum data rate of 120 Msa/s |
| M8190a | Arbitrary waveform generator with a 12-bit, 12 Gsa/s DAC and has a analog bandwidth of 5 GHz |
| PXI | Wideband signal analyzer , with a sampling rate of 2 Gsa/s |
| RSA6106A | Real-time signal analyzer with maximum analysis bandwidth of 110 MHz |
| PXA N9030 | Real-time signal analyzer with maximum analysis bandwidth of 160 MHz |
| FSW | Vector signal analyzer with maximum analysis bandwidth of 320 MHz |
| FSV | Vector signal analyzer with maximum analysis bandwidth of 40MHz |

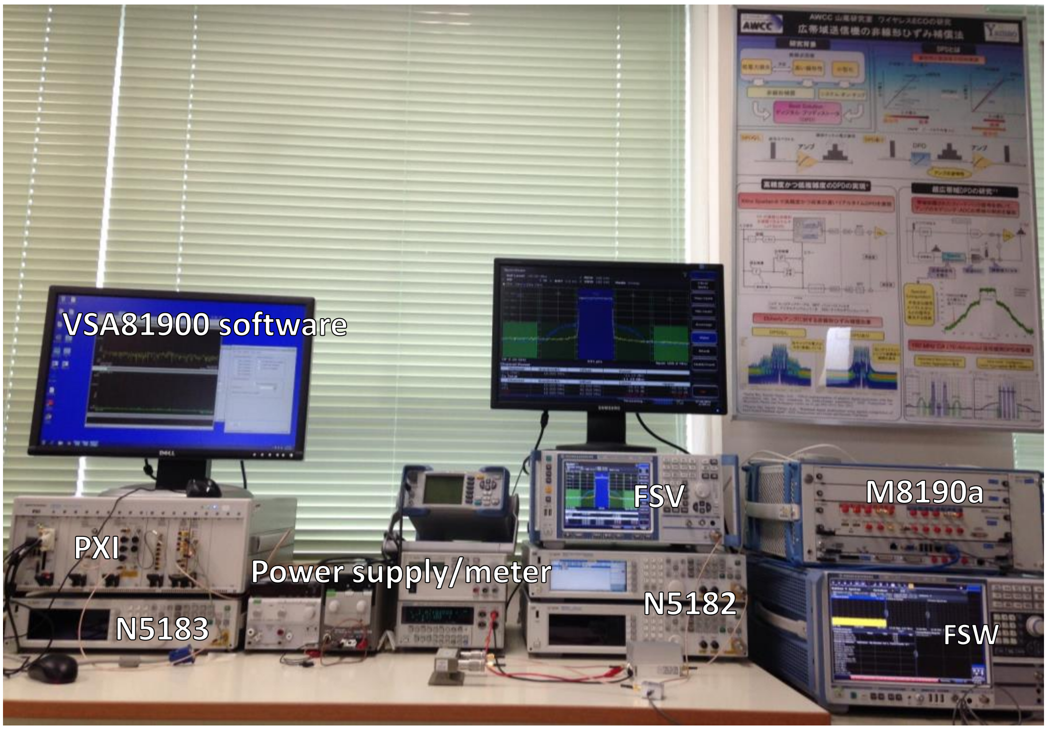


Fig. 4.3 A photo of the SG-SA test bed used in this thesis.

reference frequency for the oscillators inside the instruments, such that the frequency shift between SG and SA can be bounded under several tens Hz. There are several standards for the interface used for communications between the instruments and computer, which are GPIB (general purpose interface bus), VISA (virtual instrument software architecture), TCP/IP (local area network *abbrv.* LAN) and serial port. Can neither of the above standard provide real-time communications, hence the DPD or PNLC instrumental experiments are implemented open-loop, lacking of the implication in fast-varying conditions. The instruments are controlled by software compatible with SCPI (Standard Commands for Programmable Instruments), which is a universal syntax and commands for programming the test. There are various selections for instrument control software, including the ones from first-part instrument makers such as Agilent VSA81900, as well as third-part software such as LabView and MATLAB instrument toolbox. In this thesis, MATLAB is used as it offers much more flexibility and convenience to combine the instrument controlling program with the programs containing DPD or PNLC algorithms.

The instruments that are used in this thesis are listed in Table 4-1, and a photograph of the test bed is demonstrated in Fig. 4.3. For instance, the M8190a contains a very high-speed DAC, thereby the RF signal can be directly generated by designing the waveform in digital domain. The PXI can sample a wideband signal with bandwidth up to 1 GHz, but as its ADC has only 10 bits, the received signal has a high noise floor that may be not adequate in some circumstances. The FSW, which has a smaller bandwidth of 320 MHz and 14-bit quantization ability can provide larger dynamic range.

## 4.2 FPGA IMPLEMENTATION

FPGAs are ideal for use with high-speed peripherals, and in general it is much easier to get in hand a board that contains the necessary parts. In this thesis, the real-time implementation of DPD or PNLC was developed with FPGA evaluation board as well as FPGA mezzanine card (FMC) that contains ADC/DAC and up/down converters to form the transmission and receiving paths. A photo of the FPGA board with two FMCs are shown in Fig. 4.4 to provide a view for the test equipment we used in this thesis.

An FPGA experimental environment suitable for DPD and PNLC implementation have

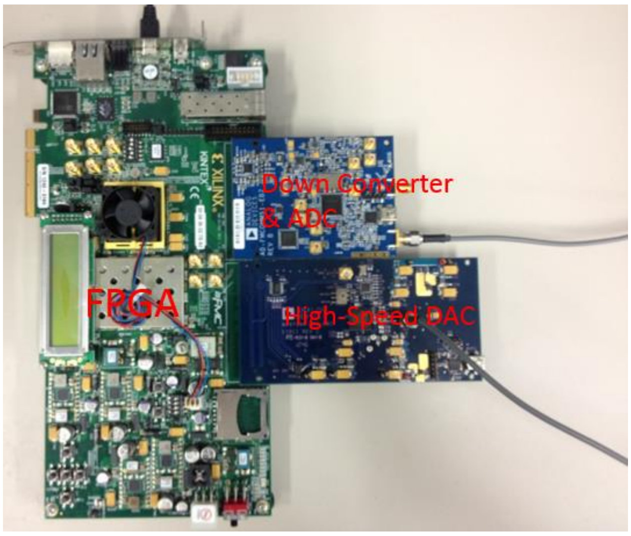


**Fig. 4.4 FPGA evaluation board and mezzanine cards used in the test.**

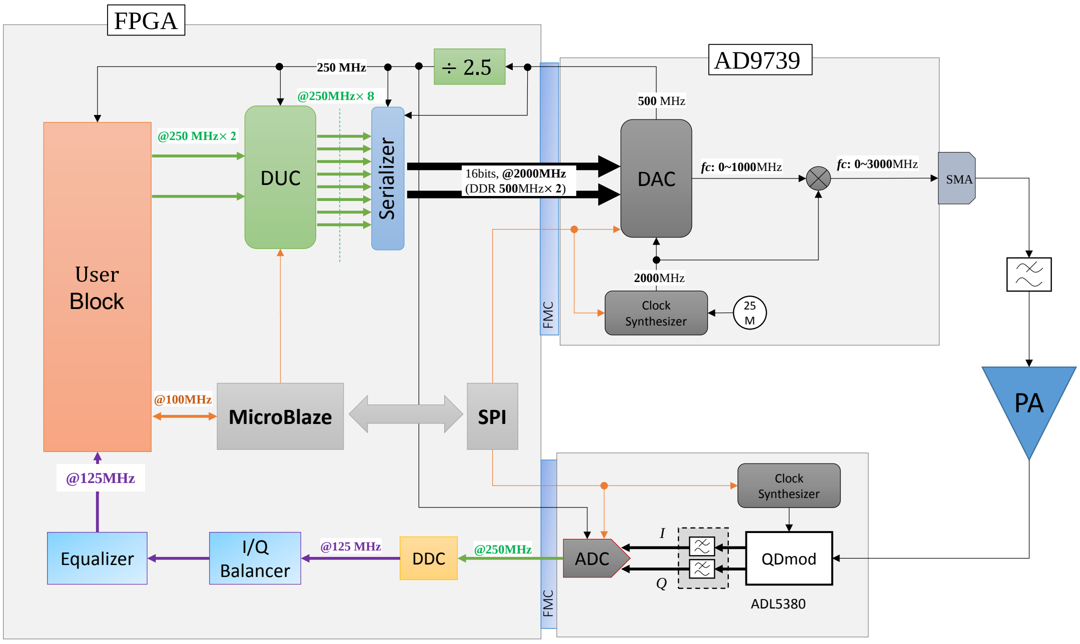


**Fig. 4.5 Diagram of the FPGA test bed used in this thesis**

been built, which is diagrammatically shown in Fig. 4.5. Two mezzanine cards, for respective transmission and receiving, were connected with the FPGA board through HPC (high pin connector) FMC interfaces, which have very high data transmission throughput. The DPD or PNLC is implemented inside the user block, which has input and output ports connected with the receiving and transmission peripherals. In addition, a simplified DSP (digital signal processor) MicroBlaze running at 100 MHz is also available to take up complicated computing tasks. The MicroBlaze also controls and monitors the peripheral chips (ADC, DAC, clock synthesizer and demodulator, etc.) with SPI (serial peripheral interface). The functional blocks in Fig. 4.5 will be introduced in the following subsections.

### 4.2.1 Transmission and Receiving Links

A 14-bit DAC with rate of 2 GSa/s (giga samples per second) is used as the transmitter, which removes the use of up converter as the DAC suffices to generate digital RF signal. A switch serving as the mixer is integrated in the DAC to extend the RF frequency coverage. Since the DAC has a data rate of 2GSa/s, the programmable transmission frequency is from 0 to 1 GHz. By using the mixer, this coverage is extended to 1~3 GHz if the DUC (digital up converter) is design to support frequency shift between -1~1 GHz. Accordingly, by setting bypass of the mixer and programming the DUC, the DAC can provide a frequency range of 0~3 GHz, which covers all the bands used for nowadays mobile communication. The DUC implemented in the FPGA is processed in parallel with 8 concurrencies, and each concurrency is under a clock of 250 MHz, thus the speed of the overall DUC is 2 GSa/s. The input to the DUC is a 250 MHz 2-paralleling signal, yielding a maximum usable bandwidth of 500

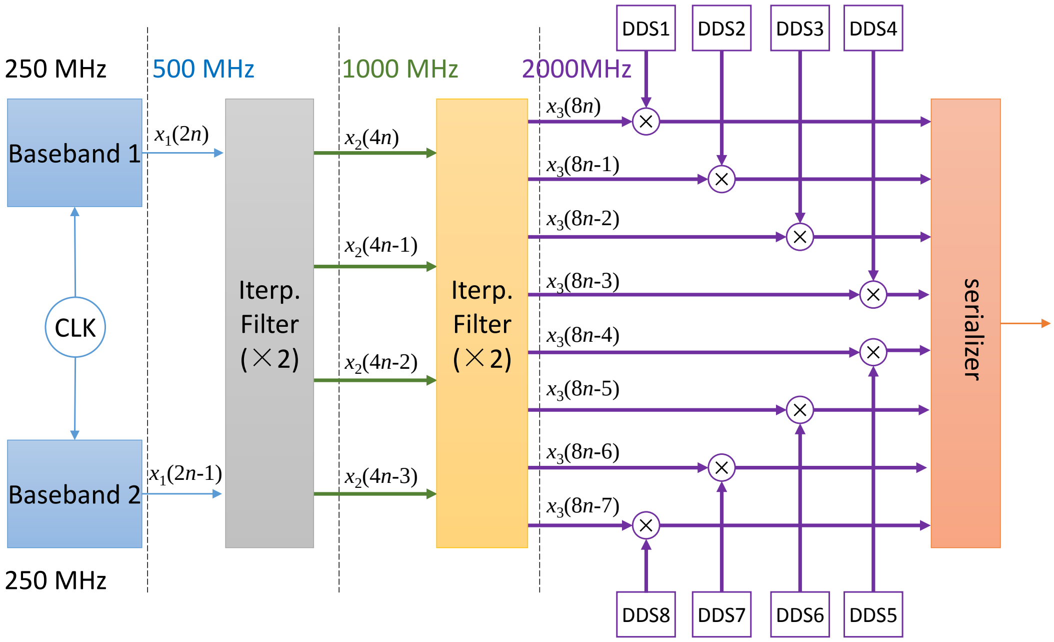


**Fig. 4.6 Diagram of the DUC**

MHz. The diagram of the DUC is shown in Fig. 4.6, from which we can see that the sampling frequency conversation is achieved by cascading of two 2× interpolation filters (half-band filters). The 8 outputs in parallel of the final stage are multiplied with a sine wave, which is generated by a DDS (direct digital synthesizer).

In the receiving path, a broadband quadrature demodulator is used to convert the RF signal to IF (intermediate frequency), and an ADC of 250 MSa/s is used to sample the IF signal. As the ADC is AC-coupled, zero-IF is not supported thereby the maximum usable bandwidth if only 125 MHz. Moreover, the I/Q impairment compensation needs not to take DC shift into account because of the AC-coupling. The digital IF signal is transferred to baseband with the DDC (digital down converter), which is followed with an equalizer to extend the distortionless usable bandwidth.

## 4.2.2 I/Q Impairment Compensation and Equalization

Contrary to the hypothetical architecture, things do not look to be perfect for IQ modulator/demodulator [100][101]. Regardless what has caused these imperfections, the signals generally suffers a gain and phase difference for I and Q channels. Fortunately, the transmission link does not contain quadrature modulator, such that the I/Q impairments existing in the receiving link can be easily identified by using a training signal or online correction.

The I/Q imbalance can be identified using the following equations:

$$y = w_1 x + w_2 x^*, \tag{4.1}$$

which is to determine the coefficients $w_1$ and $w_2$. This is a linear regressive model, so that

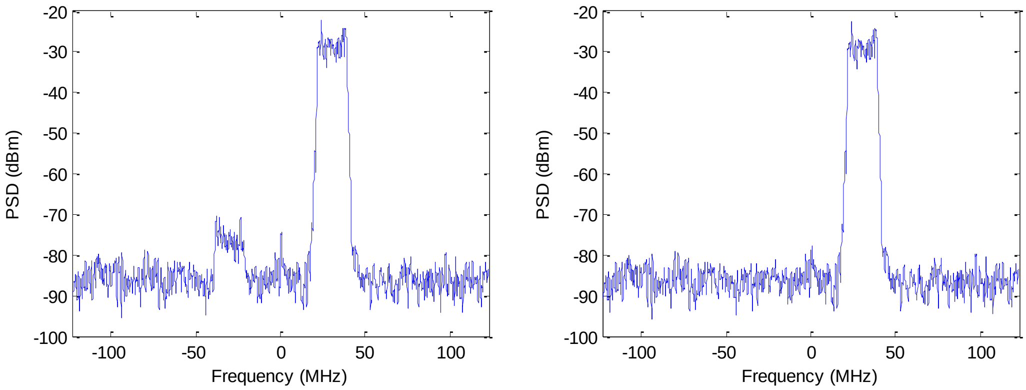


**Fig. 4.7 Digitalized IF signals without and with I/Q impairment compensation.**

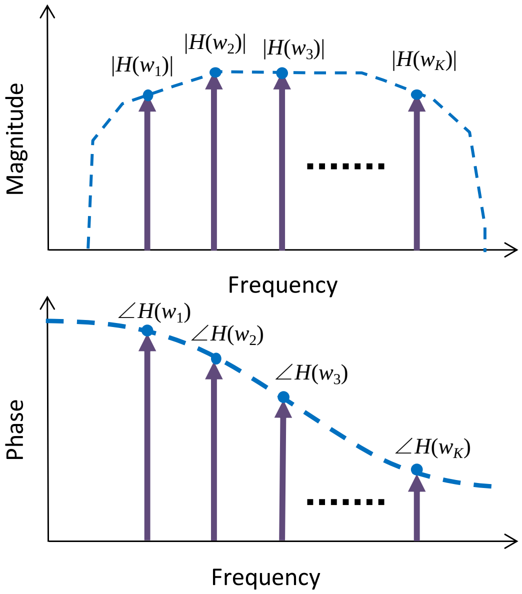


**Fig. 4.8 Use multi-tone signal to characterize the frequency response of the filter.**

the coefficients can be extracted with simple linear estimators. Fig. 4.7 shows the spectral plot of the IF signals without and with I/Q impairment compensation. One can see that the I/Q imbalance causes a mirror component in opposite band and can be totally removed after compensation.

A crucial aspect to take into account is that no extra distortions are added by the circuitry in the feedback path (including the down converter and AD converter). In our system, the anti-aliasing filter is not ideal, as it introduces linear distortion, which may be misinterpreted as the memory effect of the PA and leads to overcompensation. A commonly used method for characterizing the feedback filter is to send a single tone to sweep the bandwidth of interest, which, nevertheless, can only measure the magnitude response and lack of the

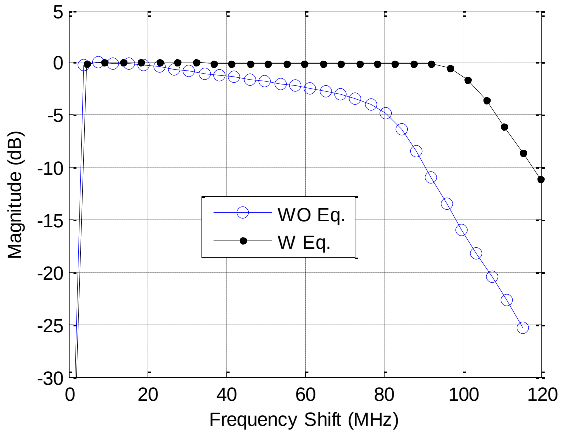


**Fig. 4.9 Magnitude-frequency response of the receiving path with and without equalizer.**

information of phase distortion. Nonlinear phase response (non-constant group delay) is another undesired distortion that is quite destructive to wideband signal inasmuch as it causes different frequency components to experience dissimilar delays. This effect hampers time alignment of the feedback signal and also transfers to additional memory effect. A simple calibration technique we have proposed in [99] is to use a *K*-tone signal, which is given by

$$m(t) = \frac{1}{K}\sum_{k=0}^{K-1} e^{j(w_k t+\theta_k)}, \tag{4.2}$$

to characterize the frequency response of the filter. $w_k$ represents the frequency shift from the center frequency, and the initial phase $\varphi_k$ can be selected to minimize the PAPR of the multi-tone signal to ensure no extra nonlinear effect to be introduced in the calibration procedure. The concept behind this method is illustrated in Fig. 4.8. Here, the output of the non-ideal filter is

$$d(t) = \frac{1}{K}\sum_{k=1}^{K-1} H(w_k) e^{j(w_k t+\theta_k)}, \tag{4.2}$$

where $H(w_k)$ is the frequency response (including amplitude and phase) for *k*-th tone. With this expression, the frequency response transfers to time domain. It is easy to see from (19) that the tones have linear response with respect to $H(w_k)$, and thus solving for the frequency response vector **H** yields a LS problem

$$\mathbf{H} = (\mathbf{M}^H\mathbf{M})^{-1}\mathbf{M}^H\mathbf{d}, \tag{4.3}$$

where **d** is a column data vector with of the output and **M** is the data matrix of the tones. With **H** in hand, the coefficients of the equalizer (an FIR filter) can be calculated by applying inverse fast Fourier transform (IFFT) to the reciprocal of **H**, and proper truncation and windowing are necessary to achieve suitable filter kernels [102].

OFDM signal can also be used for identification of the filter response, for which the procedure is similar to frequency domain equalization. The magnitude-frequency responses of the receiving path with and without equalization are plotted in Fig. 4.9. One can see that the flat frequency band is extended to 100 MHz with equalization.

### 4.2.3 Path Gain and Delay Alignment

Path gain and delay alignment is crucial in the DPD implementation. One may feel curious about why putting these two stuffs that seem irrelevant together in this subsection. The reason is simple: delay estimation based on cross-correlation is very susceptible to the magnitude difference between the two signals, especially for fractional delay.

Swing of the analog IF signal in the receiving link should be kept under the reference voltage of the data converter that defines the peak level, or otherwise overflow is to happen. In practice, one can estimate the suitable power and adjust the variable gain amplifier (VGA) to produce a signal with roughly reasonable average power inputting into the ADC. Additionally, the filtering and frequency conversion in digital domain may involves truncating of the LSB (least significant bit) and necessary bit shifting, thereby the resulting signal may

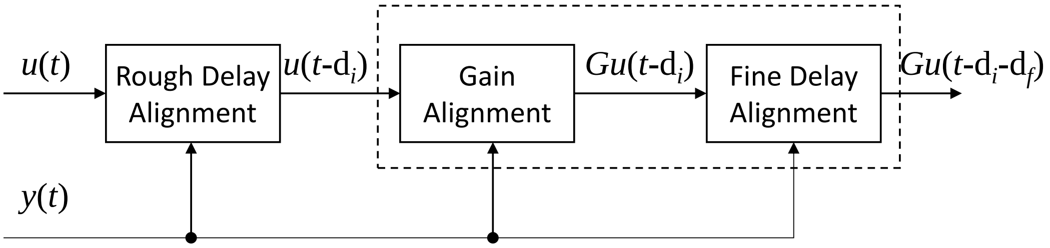


**Fig. 4.10 The flow of time and gain alignment.**

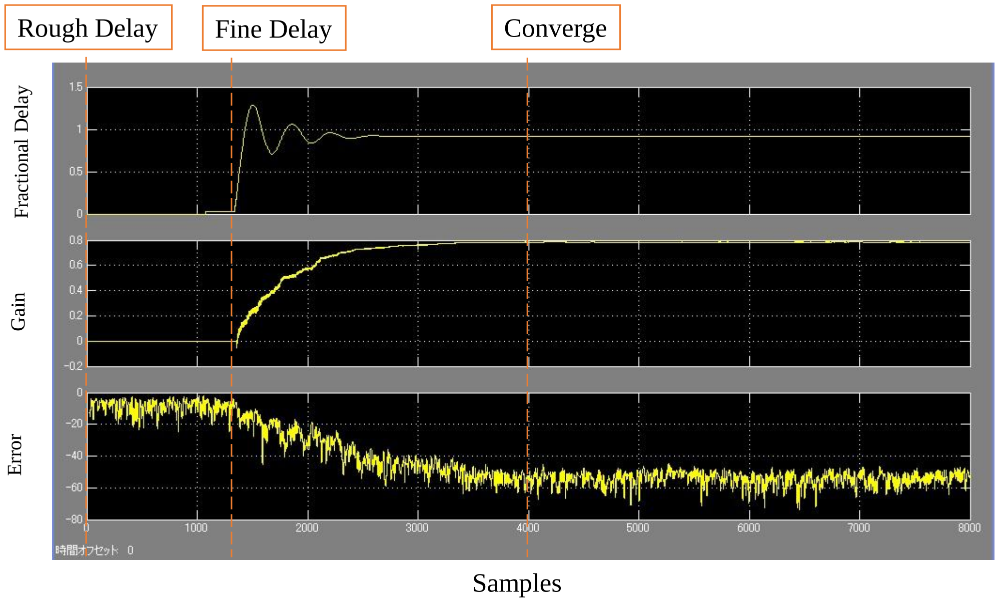


**Fig. 4.11 The hardware co-simulation results for time and gain alignment.**

have much lower power than the intact signal.

The delay alignment is carried out by taking integer delay estimation that followed by fractional delay estimation. The integer delay estimation can be done by correlating the two signals, and find the peak of the correlation function, while the fractional delay is accomplished with interpolation filter, such as Farrow filter [103]-[106].

The flow for gain and time alignment is depicted in Fig. 4.10, which from left to right contains a rough (integer) delay alignment, gain alignment and fine (fractional) delay alignment. The gain and fine delay alignments are taken place when rough delay alignment is done and they are executed simultaneously. $u(t)$ denotes the signal to align and $y(t)$ is the reference signal. The gain and fractional delay estimation can utilize linear estimators, such as the LMS (least mean squares) that is simple to implement. From the hardware co-simulation results taken place in MATLAB Simulink/System Generator that shown in Fig. 4.11, one can see that the rough time alignment takes about 1200 samples, and then the fine delay and gain alignment is triggered and come to stable status after 4000 samples. The estimated fractional delay is 0.9 UI (unit interval), and the gain difference is 0.8.

# Chapter 5.

# FPGA Implementation of Real-Time Digital Predistortion with Low Complexity and High Convergence Rate

This chapter represents an adaptive digital predistorter (DPD) with fast convergence rate and low complexity for multi-channel transmitters, which is fully implemented in an FPGA (field programmable gate array). The design methodology and practical implementation issues are discussed, with concerns about the impact caused by carrier power shutdown and transmission power control. The proposed DPD is composed of multiple adaptive lookup table (LUT) units of uniform structures, allowing configurability for desired memory depth. A simplified multiplier-free normalized least mean square (NLMS) algorithm for fast adapting the LUT is introduced. The proposed DPD is also experimentally exploited to linearize a Doherty amplifier. The adjacent channel leakage ratio (ACLR) reaches -60 dB, for both lower and upper bands, in the test applying long term evolution (LTE) signal. It is also demonstrated in this chapter that the proposed DPD shows high robustness when multi-channel GSM (global system for mobile communications) signal with occasional carrier power shutdown is applied.

The organization of this chapter is as follows. Section 5.1 presents the effect of signal variation on the PA behavior, which justifies the use of real-time high convergence rate DPD. Section 5.2 discusses the differences between polynomial and LUT ahead of the description of the DPD model and algorithm. Section 5.3 presents the detailed issues about the implementation of proposed DPD in the FPGA. Experimental results are shown in Section 5.4, followed by a brief conclusion in Section 5.5.

## 5.1 THE NEED FOR REAL-TIME DPD WITH LOW COMPLEXITY

It has been taken for granted by many researchers that the DPD does not need to be consecutively adapted once the DPD coefficient set is found, owing to the slowly time-variant property (device aging and temperature drifting) of the transistors. Such DPD is referred to as *open-loop* [20][107] or slow-adaptive. The well-established least squares (LS) algorithm

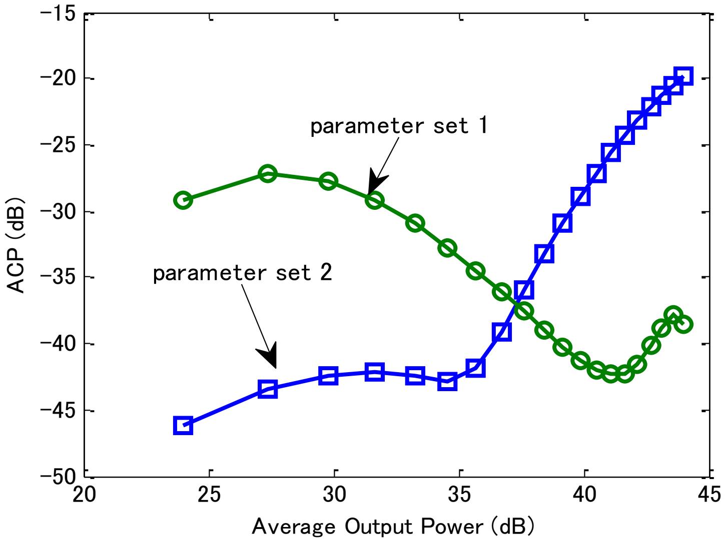


**Fig. 5.1 The measured ACP of a real PA output applying two different DPD parameter sets extracted at different power level employing least square fitting.**

is widely applied for *open-loop* DPD parameter estimation due to its robustness and minimized quadratic error in deterministic sense. Albeit the slowly varying physical conditions of the PA, the amplifiers in the commercial transmitter are time variant seen from the baseband, mainly due to the fact that the PA has dissimilar characteristics with different exciting signals[108][109]. Thus, while adequate for the steady signals, the open-loop or slow-adaptive DPDs fall short of rapid response to the signal property changes.

The first factor that should be taken into account is that the DPD correction performance depends on the average input power variation after one set of parameters is updated to the DPD. The power variation refers to the transmitted average power change due to the traffic load condition variation; this is precisely what happens in the presence of call-load changes and power control in a cell site. The effect caused by power variation on the DPD is reported in Fig.5.1 where two sets of DPD parameters are extracted employing least squares. With the parameter set 1, which is calculated using the signal with average power being set at a higher power level, the adjacent channel power (ACP) significantly degrades as the average output power varies from the optimum point. The second case, on the other hand, shows optimized performance at a lower power level at where the parameters were extracted.

The signal variation phenomenon becomes even more destructive for multi-channel transmitter in which different users are distributed on respective carriers that are occasionally shut down to save energy. This feature of the signal adds another dimension of variation and has a significant impact upon DPD design, since not only the average power, but also the probability distribution, peak-to-average power ratio (PAPR) and even bandwidth vary with time. This represents a far more unwieldy hurdle for open-loop and slow-adaptive DPDs.

In some particular implementations, such as the repeater, the input signal variation is unknown beforehand. This causes severe problems and hinders standard-compliant operation. For instance, when all carriers are suddenly shut down, the input signal power at the repeater side becomes considerably low (the input signal becomes noise) at this incident. Such occurrence may result in DPD corruption if the adaption fails to pause in time.

Drawing inspirations from the foregoing discussions, the key requirement, stressed before, poses the main challenges when designing a predistorter: the DPD should have fast adaption rate in order to track the characteristic variation of the PA (mainly caused by signal variation). Secondly, the DPD is also required to be robust to sudden incidents (such as the carrier power shutdown events). In addition, the DPD should be less susceptible to the signal properties to enable all-standard -compliant operation.

An adaptive real-time LUT predistorter with low complexity and fast convergence rate for multi-channel transmitters is reported in this chapter herein the design art and practical FPGA implementation issues are presented. The DPD has high reconfigurability as which is constructed with multiple adaptive LUT units of uniform structures. A suboptimal multiplier-free algorithm of low-complexity is proposed to adaptively adjust the DPD. The effect of phase shift is also analyzed in this chapter, as it has a significant impact on the convergence of the direct learning DPD as the one presented in this chapter. The proposed DPD is advantageous for many reasons, not the least of which is its indifference towards signal variation.

## 5.2 DPD ALGORITHM AND ARCHITECTURE FOR FPGA IMPLEMENTATION

Before delving into the details of the LUT predistorter implementation, it is instructive to clarify the differences between the LUT and polynomial DPDs. In the following discussion, the appealing advantages of LUT DPD over its counterpart are highlighted. To start with, LUT has much lower complexity than polynomial in terms of fewer multipliers. Secondly, subjecting to the precession of the fixed-point digital signal processor (DSP), the polynomial has potential numerical stability problems. As a rule of thumb, a 32-bit processor can only handle polynomials no more than 5th or 6th order, if no extra numeric stable algorithm is exploited. Thus, LUT predistorter is a natural consequence of practical fixed-point implementation, especially for FPGA [110][111]. Moreover, the polynomial-based adaptive predistorter is less robust to quick power variations, in other words, such DPDs should be notified to stop adaption when the power is violently fluctuating, whereas this is always difficult to be ensured. For instance, sudden shutdown of the transmitted signal may cause the polynomial adaption to corrupt (diverge), inasmuch as the input signal becomes white noise at such times. Commercial DPD always applies an extra block to monitor the signal's quality (like the probability density function) and decide whether the DPD should be updated. The LUT predistorter, on the contrary, is inherently immune to this problem given that each address of the LUT is updated independently. When signal shutdown occurs, the input signal level becomes considerably low whereby only the first or first several address/addresses of the LUT is/are visited. This will not cause a severe problem because a few wrong-updated addresses cause limited performance degradation. In addition, the LUT is also less susceptible to PA characteristic change caused by signal power variation, as the LUT is a function segmented by input power.

### 5.2.1 LUT DPD Model

The DPD model applied here is a combination of $M$ paralleled-connected LUTs as delineated in Fig.5.2. This model is essentially a manifestation of the memory polynomial, which is one of the most popular models for compensation of PAs that represent moderate nonlinearity and memory effect. Its expression is given as

$$x(n)=\sum_{m=1}^{M}\sum_{i=0}^{N-1}a_{ij}\left|u(n-m+1)\right|^{i}u(n-m+1), \tag{5.1}$$

where $N$ is the nonlinear order, and $M$ is the memory depth. By carefully examining this function, it can be found that the memory polynomial is constituted by $M$ paralleled memoryless polynomials, as shown below

$$\begin{aligned} x(n)=&\sum_{i=0}^{N-1}a_{i1}\left|u(n)\right|^{i}u(n)+\sum_{i=0}^{N-1}a_{i2}\left|u(n-1)\right|^{i}u(n-1)+\\ &\cdots+\sum_{i=0}^{N-1}a_{iM}\left|u(n-M+1)\right|^{i}u(n-M+1), \end{aligned} \tag{5.2}$$

which yields a reasonable expression that is suitable for FPGA implementation

$$\begin{aligned} x(n)=&G_{1}\left(\left|u(n)\right|\right)\ u(n)+G_{2}\left(\left|u(n-1)\right|\right)\ u(n-1)+\cdots\\ &+G_{M}\left(\left|u(n-M+1)\right|\right)\ u(n-M+1), \end{aligned} \tag{5.3}$$

with $G_m(|u(n-m+1)|)$, $m= 1, 2, \ldots M$, being the complex gain. It is easy to see from (5.3) that each of the memoryless polynomials can be replaced by the multiplication of the signal and

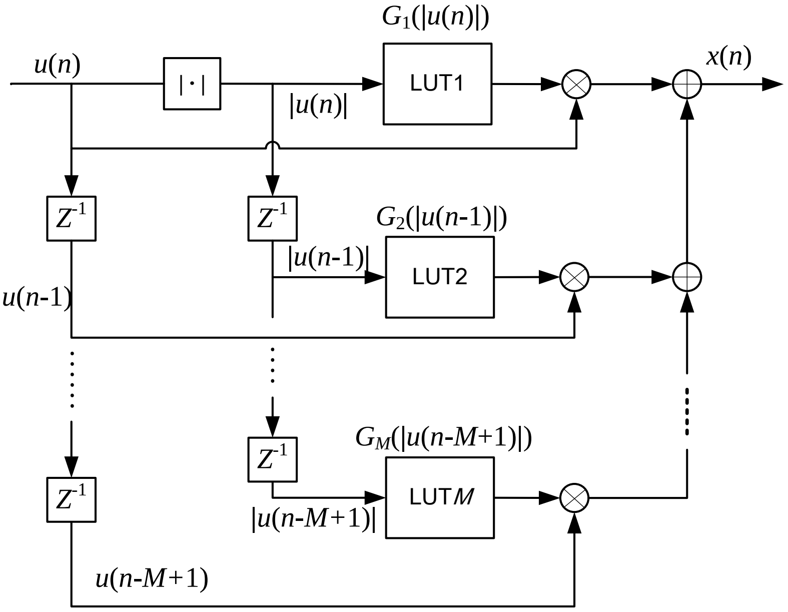


**Fig. 5.2 LUT-based predistorter model.**

an LUT indexed by the signal amplitude, hence the memory polynomial can be seen as a combination of several LUTs, as shown in Fig.5.2. Such a model is easy for FPGA implementation using fixed-point arithmetic as there are only $M$ complex multipliers and one modulus operator involved.

### 5.2.2 Multiplier-Free NLMS Adaption Algorithm

The updating method used here is the *direct learning*, which just as its name implies, directly estimates the pre-inverse of the PA. This forms a closed-loop structure so that the DPD is updated consecutively, *viz.* in a sample-by-sample manner, with aim to minimize the error between the input signal and the attenuated PA output.

Associating with the LUT units, $M$ adaption blocks sharing the same error function are exploited, as demonstrated in Fig.5.3, to independently update the LUTs. Because of the uniform structure, the model configuration, more specifically the memory depth $M$, can be easily altered according to the memory effect intensity of the target PA.

To derive the adaption function for the LUT, we first consider a single LUT in upstream of the PA, and the respective signals are given as

$$\begin{aligned} v(n) &= G\left(|u(n)|\right)u(n) \\ y(n) &= f\left(v(n)\right) = f\left(G\left(|u(n)|\right)u(n)\right), \end{aligned} \tag{5.4}$$

where $G(\ )$ is the content of the LUT and $f(\ )$ denotes the function of the PA. The objective of DPD is to minimize

$$J = E\left\{\left|u(n) - \frac{y(n)}{G_{lin}}\right|^2\right\} = E\left\{\left|u(n) - \frac{f\left(G\,u(n)\right)}{G_{lin}}\right|^2\right\}, \tag{5.5}$$

where $G_{lin}$ is the desired linear gain, which should be controlled to prevent overloading the PA, and the parenthesis behind $G$ is omitted for conciseness. Without loss of generality, $G_{lin}$ is set to 1 for simplicity, and hence (5.5) yields to

$$\begin{aligned} J = E\{&u^*(n)u(n) - u^*(n) f\left(G\,u(n)\right) - u(n) f^*\left(G\,u(n)\right) + \\ &f\left(G\,u(n)\right) f^*\left(G\,u(n)\right)\}\ . \end{aligned} \tag{5.6}$$

To find the minimum of $J$, it always requires computing the first-order differentiation of $J$ with respect to $G$. By inspecting (5.6), however, we will notice that $f(\ )$ is an unknown function so that the differentiation cannot be explicitly derived. Therefore, the predistorter is a nonlinear minimum estimator[2] that does not have an analytic form in general. Consider that the PA has weak nonlinearity, more specifically, the intermodulation distortion components

---

[2] The indirect learning DPD is a linear estimator as it estimates the post-inverse, which is linear in respective to its parameters, instead of directly estimating the pre-inverse, as already discussed in Section 3.4.1.

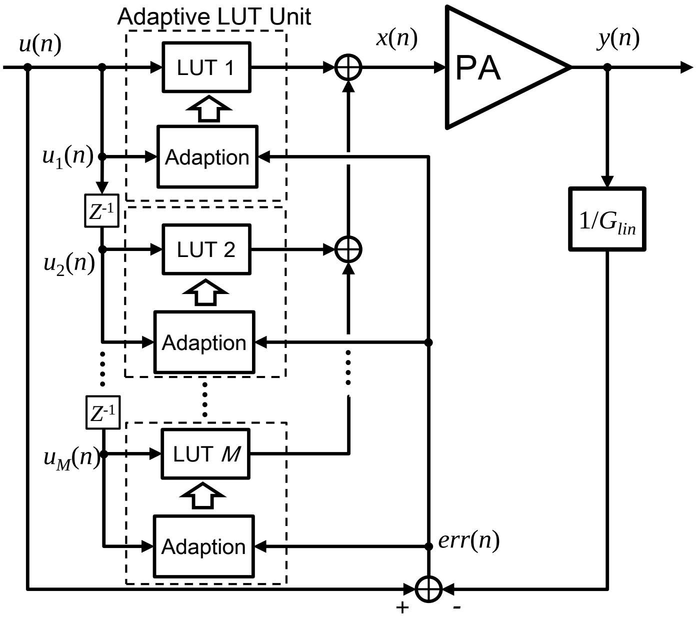


**Fig. 5.3 Simplified block diagram of the predistortion system, where the DPD is constituted of *M* independent adaptive LUT unit**

have much lower power (e.g. several tens of decibel) than the linear component, *f*( ) can be approximated with a linear function so that

$$J \approx \tilde{J} = E\left\{u^*(n)u(n) - u^*(n)G\,u(n) - G^*u^*(n)u(n) + GG^*u^*(n)u(n)\right\} \;. \tag{5.7}$$

The differentiation of (5.7) becomes

$$\begin{aligned} \frac{\partial \tilde{J}}{\partial G^*} &= 2E\left\{Gu(n)u^*(n) - u^*(n)u(n)\right\} \\ &= 2E\left\{u^*(n)\left[Gu(n) - u(n)\right]\right\} \\ &\approx 2E\left\{u^*(n)\left[y(n) - u(n)\right]\right\} \;, \end{aligned} \tag{5.8}$$

with which the least mean square (LMS) can be derived:

$$G_{i+1} = G_i + \gamma u^*(n)\left[u(n) - y(n)\right] \;, \tag{5.9}$$

where $\gamma$ is the step size. Unlike the linear optimization problem, there requires some prerequisites for (5.9), such as small phase shift between $u(n)$ and $y(n)$, to ensure the convergence of the DPD. This is can be understood by adding a constant phase shift $\varphi$, and the target of predistortion is to

$$f\left(Gu(n)\right)\exp(j\phi) \overset{set}{=} u(n) \;. \tag{5.10}$$

The solution of $G$ becomes

$$G' = \frac{f^{-1}\left(u(n)\exp(-j\phi)\right)}{u(n)}. \tag{5.11}$$

If *f*( ) is a linear function, the constant phase will only rotate the estimated *G* from the desired solution with $-\varphi$. Whereas, nonlinear *f*( ) will lead to wrong estimation, as will be discussed in Section 5.3.3. In practice, the constant phase shift is not necessarily equal to 0, but it should be as small as possible.

Applying (5.9) to the DPD shown in Fig.5.3, the update equation for *m*-th LUT unit for (*i*+1)-th iteration is given as

$$G_m(\bullet)_{i+1} = G_m(\bullet)_i + \gamma u^*(n-m+1)\left[u(n) - \frac{y(n)}{G_{lin}}\right]. \tag{5.12}$$

The addresses in the parentheses are replaced by • for simplicity.

By examining (5.12), it is apparent that LMS performs unfair update for different entries: the upper entries converge faster than lower ones. To hasten the convergence, a normalized LMS (NLMS) can be applied

$$G_m(\bullet)_{i+1} = G_m(\bullet)_i + \gamma \frac{u^*(n-m+1)}{\left|u(n-m+1)\right|^2}\left[u(n) - \frac{y(n)}{G_{lin}}\right]. \tag{5.13}$$

The NLMS, however, is not an efficient implementation for FPGA. The reasons are twofold: firstly, divider is an expensive operation that is not amenable to efficient hardware implementation; secondly, NLMS is susceptible to small perturbations, potentially yielding numerical instability. To make the implementation feasible in hardware while maintaining the fast convergence feature, the NLMS is approximated by

$$G_m(\bullet)_{i+1} = G_j(\bullet)_i + \gamma \frac{u^*(n-m+1)}{\left|u(n-m+1)\right|}\left[u(n) - \frac{y(n)}{G_{lin}}\right], \tag{5.14}$$

with which all the addresses can be updated with equal rate. The appealing hallmark of this expression is that it can be reformulated as

$$G_m(\bullet)_{i+1} = G_m(\bullet)_i + \gamma \exp\left(-j\theta(n-m+1)\right)\left[u(n) - \frac{y(n)}{G_{lin}}\right], \tag{5.15}$$

where $\theta(n\text{-}m+1)$ is the instantaneous phase of $u(n\text{-}m+1)$. The divider is avoided, and the multiplication of $\exp(-j\theta)$ can be accomplished by vector rotation, which is numerically superior to division.

The computational load of the vector rotation can be further reduced by quantizing the phase $\theta(n\text{-}m+1)$. In our implementation, 1-bit quantization is made:

$$\begin{aligned} G_m(\bullet)_{i+1} = G_m(\bullet)_i + \gamma \cdot \big[ & sign\left(u_{real}(n-m+1)\right) - \\ & j \cdot sign\left(u_{imag}(n-m+1)\right)\big] \cdot \left(u(n) - \frac{y(n)}{G_{lin}}\right). \end{aligned} \tag{5.16}$$

Considered that the gradient of the one-tap learning is dominated by the error vector, quantizing phase of $u$ does not create obvious bottleneck on the performance [112].

### 5.2.3 LUT Spacing

It has been shown by J. K. Cavers that the DPD performance can be significantly improved by using optimal address spacing [113] that can be deduced when the amplifier AM-AM and AM-PM characteristics are known *a priori.* However, it has also been shown in [113] that a few decibels of backoff exacerbates the performance of initial optimal spacing. Therefore, maintaining optimal performance across all power levels require redefining the spacing once the signal average power varies, which is infeasible in practical implementations. The effect of power variation on the LUT performance is obvious, especially when the signal power level is low: smaller portions of the LUT addresses are visited, which results in less effective bits and lower signal to noise ratio (SNR).

From the foregoing discussion, the question arises as to what is the spacing method can be used to alleviate SNR reduction against power variation. An applicable way is to exploit the well-established $\mu$-law algorithm [114] to remap the addresses so that the performance can be balanced in a wide power range. The $\mu$-law compander is given as

$$F(|u|) = \frac{\ln(1+\mu|u|)}{\ln(1+\mu)}, \tag{5.17}$$

where $\mu$ can be adjusted according to different applications.

## 5.3 FPGA IMPLEMENTATION ISSUES

The block diagram of the LUT DPD system is demonstrated in Fig. 5.4. In transmit path, the baseband I/Q signal is converted to intermediate frequency (IF) with a digital up converter (DUC) block that contains interpolators and a frequency shifter. The feedback signal is digitized by the ADC and transferred to baseband with a down converter (DDC). The equalizer is exploited to remove the linear distortion caused by the anti-aliasing filter. The input signal should be properly time aligned and phase aligned with the feedback signal to allow correct calculation of the error. The constant phase shift compensation is compulsory for the direct learning DPD, which is susceptible to the phase difference. The instantaneous error is calculated by subtracting the feedback signal from $u'(t-\tau)$, which is the time and phase aligned version of the input signal $u(t)$. The $\mu$-law function, which is stored in a block RAM, produces the address for the LUT units. Implementation of the adaptive LUT units, loop delay adjuster and constant phase shift adjuster will be discussed in detail hereinafter.

### 5.3.1 Implementation of the Adaptive LUT Unit

The adaptive LUT unit can be duplicated and combined in parallel to form the adaptive memory polynomial DPD of desired memory depth, due to its uniform structure. The $m$-th

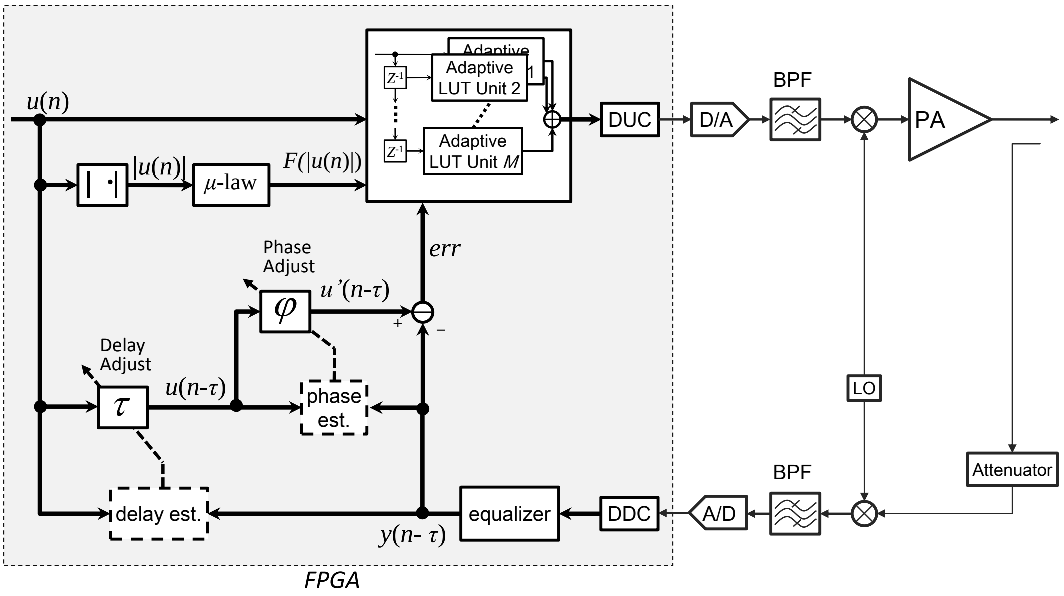


Fig. 5.4 Implementation of digital predistortion system with FPGA.

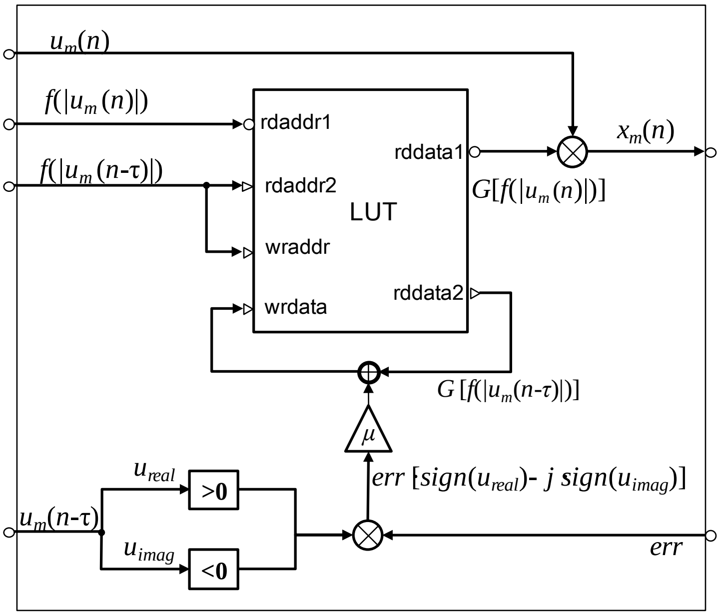


Fig. 5.5 Block diagram of the ***m***-th adaptive LUT unit.

adaptive LUT unit is illustrated in Fig.5.5, which contains an LUT and the simplified NLMS block. The output signal $x_m(n)$ is the product of the input signal for $m$-th LUT unit $u_m(n)$ (equivalently $u(n\text{-}m\text{+}1)$) and the corresponding complex gain. The adaption requires to use the delayed signal $u_m(t\text{-}\tau)$, as well as the LUT output at the instance of $t\text{-}\tau$, since the feedback signal lags behind the input signal by $\tau$ unit intervals (UIs). Moreover, the address of the LUT that is to be updated should also be addressed by the amplitude of $u_m(n\text{-}\tau)$ to conform the time delay. Based on the above discussion, it is clear that some delay-tunable registers

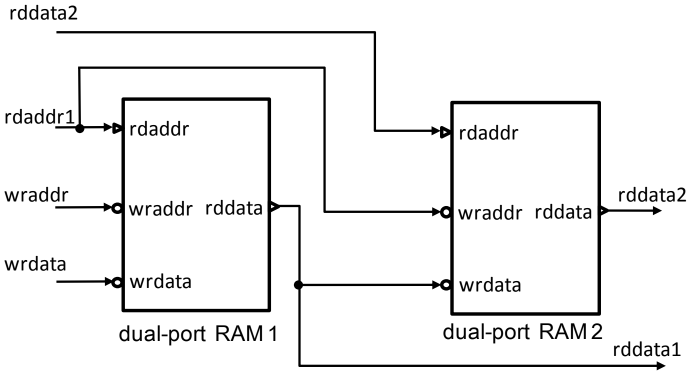


**Fig. 5.6 Block diagram of the *m*-th adaptive LUT unit.**

are necessary to align these signals. However, this hinders the reconfigurability of the DPD. As a result, a triple-port RAM that has two output ports (read address/data port) and one input port (write address/data port) is used as the LUT to avoid the use of such registers.

According to (5.15), the error is rotated by -$\theta(n-m+1)$, which is obtained by estimating the instantaneous phase of $u'(t-\tau)$. Two coordinate rotational digital computer (CORDIC) cores [115], in translate and rotation mode, respectively, are required to implement this equation. Instead, the equation (5.16) offers a much simpler scheme to implement the NLMS. The real and imaginary parts of $u_m(t-\tau)$ are quantized to 1-bit and then multiplied with the error. Actually, the multiplier hereof can be replaced by several inverters, yielding a multiplier-free implementation. The step size $\gamma$, which is used to coordinate the convergence rate, can be set to 2, 1, 0.5, 0.25, etc. to avoid the use of multipliers: left and right bitwise shifting are equivalent operations.

Two dual-port block RAMs can be used to construct the LUT unit, which is a triple-port RAM as is depicted in Fig.5.6. A dual-port RAM has one read address/data port and one write address/data port, respectively, to enable simultaneous read and write operation. The read address of first dual-port RAM (denoted by RAM 1) is directly connected to the write address of the second RAM, and the output (read data) of RAM 1 is fed to the write data port of the RAM 2, allowing data sharing between the two RAMs. Once the port rdaddr1, which corresponds to the address for the gain needed for calculation of predistorted signal (see Fig.5.5), is addressed, the output data is copied to RAM2. Since rdaddr1 is always in advance of rdaddr2, the corresponding content has been updated before it is addressed, ensuring no ambiguity to be created in the reading and writing procedures.

Selecting the LUT size is another important aspect of DPD implementation, as which dominates compromise between accuracy and convergence rate. Normally, if we assume the convergence time for a single LUT to be $T_c$, the model in Fig.5.2, which is constituted by $M$ LUTs with the same size, will converge in approximately $T_c^M$, considering that the LUTs are independently updated. In order to achieve faster convergence rate, the LUT size should be decreased, while the accuracy is sacrificed. As will be shown in Section 5.4.1, the first LUT in the model dominates the accuracy, and decreasing the size of the other LUT units can increase the convergence with negligible loss of accuracy.

## 5.3.2 Loop Delay Adjuster

The loop delay compensation is constructed by two blocks: delay estimation and adjustment. The delay adjustment can be easily accomplished by utilizing a dual-port RAM and keeping the read port addressed by a counter lagging behind another counter that addresses the write port by the desired delay value.

The basic idea of estimation of the delay is to correlate the two signals, and find the delay value so that the correlation is maximized. However, directly computing the cross correlation is an exhaustive work. An often-used method is to apply the following equation

$$\arg : \max_{\tau} \left\{ \sum_{n=1}^{L} D\left[u(n)\right] D\left[y(n-\tau)\right] \right\} , \tag{5.18}$$

where $L$ is the sequence length that to calculate the correlation, and

$$D\left[u(n)\right] = sign\left[\left|u(n)\right| - \left|u(n-1)\right|\right] \tag{5.19}$$

is the sign of the amplitude difference. Since $D[\cdot]$ is single bit, the multipliers in (5.19) degrade to *XNOR* (exclusive *NOR*) gates.

Fig. 5.7 demonstrates a delay estimator that can search the delay ranging from 0 to $M$-1 UIs. The amplitude differences of the input signal and feedback signal, both in single bit, are inputted to two FIFO (first in first out) registers with size of $L$. When the free-running $M$-counter's output becomes zero, the contents of the FIFO containing the feedback signal are copied to a register. The input signal that stored in the FIFO and the feedback signal in the register are fed to an array of *XNOR* gates, and a fully pipelined bit counter that counts the number of 1's is employed to calculate the cross correlation. Since the counter counts back to 0 when the maximum value $M$-1 is reached, the register renews the data every $M$-UIs. The FIFO, on the other hand, delays the input signal one UI when a new data comes, so that the

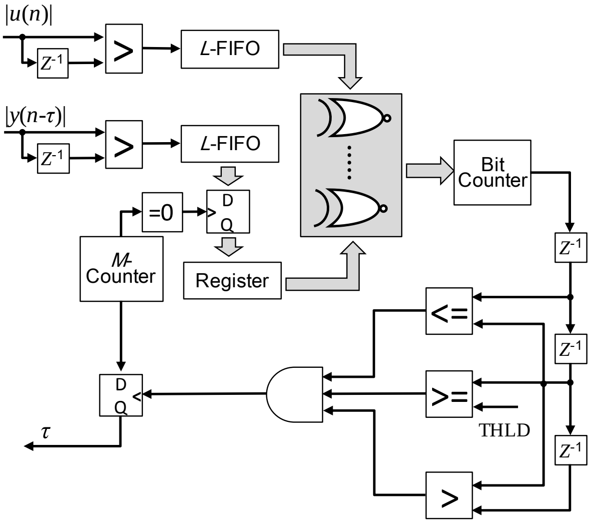


**Fig. 5.7 Implementation of the delay estimator, which searches the delay from 0 to *M*-1 UIs.**

value of the $M$-counter corresponds to the delay when the maximum correlation is found. The THLD in the figure denotes a preset threshold value, and the peaks higher than the threshold will be recognized as the maximum point. As the magnitude of the peak of the correlation coefficients depends on the extra signal applied, it generally requires an offline test to find the proper value for THLD.

It is worth to mention that (5.19) involves the calculation of the amplitude of complex number given the real and imaginary part, i.e., the square root of the sum of two squares, thus there normally requires an extra CORDIC block. Nevertheless, the calculation of the amplitude for $D[\cdot]$ is not necessary to be so accurate, and hence a so called $\alpha$Max + $\beta$Min method [116] can be exploited to estimate the approximation of the absolute amplitude of the complex number. The approximation, just as its name implies, is expressed as

$$\sqrt{I^2+Q^2}=\alpha\max\{|I|,|Q|\}+\beta\min\{|I|,|Q|\}\ . \tag{5.20}$$

For most precise estimation, the optimum values [36] of $\alpha$ and $\beta$ are

$$\alpha=\frac{2\cos\frac{\pi}{8}}{1+\cos\frac{\pi}{8}}\approx 0.9604 \qquad \beta=\frac{2\sin\frac{\pi}{8}}{1+\cos\frac{\pi}{8}}\approx 0.3978. \tag{5.21}$$

In this case, nevertheless, two multipliers are involved.

In order to further reduce the complexity for implementing $\alpha$Max + $\beta$Min while preserving acceptable accuracy, some smart choices of $\alpha$ and $\beta$ are more desirable. One pair of such parameters is $\alpha$ = 1 and $\beta$ = 3/8, and this will give a maximum error of 3.97%. With such parameters, the multiplications are reduced to some simple operations, see Fig. 5.8. The “>>1”, for instance, denotes a bitwise rightward shift by 1 bit, and which is equivalent to a division-by-2 operation. The comparator, which is denoted by >, outputs TRUE when $|I|$ is

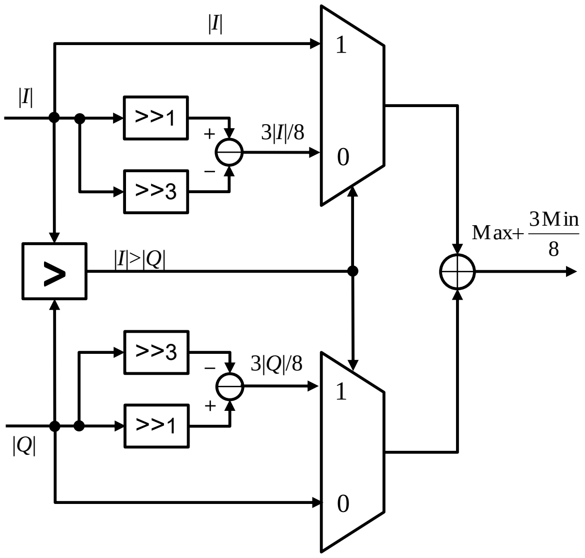


**Fig. 5.8 Implementation of a amplitude estimator, where $\alpha$=1 and $\beta$=3/8.**

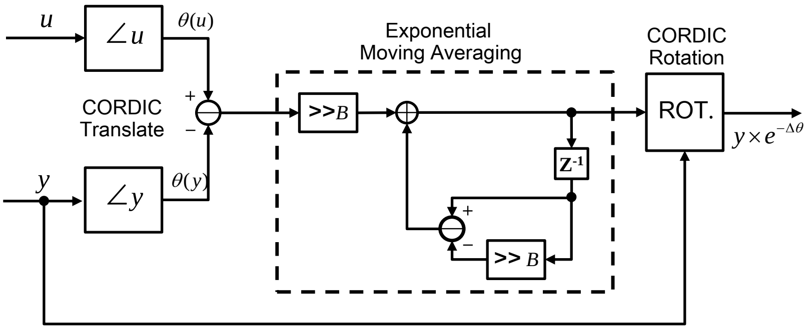


**Fig. 5.9 Implementation of the constant phase shift adjuster.**

greater than $|Q|$, and vice versa. This signal controls the multiplexers to produce $|I|+3|Q|/8$ or $|Q|+3|I|/8$ after the adder.

## 5.3.3 Constant Phase Shift Adjuster

In addition to the amplitude–dependent phase conversion (AM-PM) effect introduced by the PA, the peripheral circuits as well as the PA matching network also cause a constant phase shift between the feedback signal and the input signal. For indirect learning DPD, the effect of this phase shift can be neglected as the post-inverse is a linear estimator. However, for direct learning DPDs, as the one used here, are susceptible to this phase shift, as already discussed before. Therefore, the phase shift should be kept small, which mandates the use of an extra phase compensator.

To implement the phase adjustment, as shown in Fig.5.9, two CORDIC cores working in vector translate mode are used to estimate the instantaneous phases for the input signal $u$ and feedback signal $y$, respectively. In vector translate mode, the complex signal in rectangular form is rotated using CORDIC algorithm, until the imaginary part becomes zero. The calculated phase is fed to another CORDIC in rotation mode. The constant phase shift is approximated by the mean phase difference which is estimated with an exponential moving averaging [117] block. The expression representing the exponential moving averaging is

$$v_{ema}(n) = \delta v(n) + (1-\delta) v_{ema}(n-1). \tag{5.22}$$

where $\delta$ is the weighting factor which equals to $1/2^B$ in the Fig.5.9.

## 5.4 EXPERIMENTAL RESULTS

The device under test (DUT) is a Doherty PA, which is an LDMOS (laterally diffused metal oxide semiconductor) amplifier using MDE6IC9120N from Freescale [118]. Its match design is optimized at 945 MHz. A two-stage driving amplifier, which provides a gain of 35 dB, was used to drive the Doherty amplifier. Fig. 5.10 shows the measured gain and power-added efficiency (PAE) of the Doherty amplifier versus output power at 945 MHz. It can be found that the maximum output power is around 51 dBm.

The total FPGA utilization is summarized in Table I, where the resource consumptions

**Table 5-1 FPGA Utilization Summaries**

| Module | Slice Registers | Slice LUTs | Block RAM/FIFO | DSP48A1s (Multipliers) |
|---|---|---|---|---|
| DPD (2 adaptive LUT units and $\mu$-law) | 3,051 | 2,103 | 5 | 6 |
| Rough Time Alignment | 3,244 | 2,311 | 2 | 0 |
| Fine Time Alignment | 2,596 | 2,646 | 3 | 5 |
| Magnitude Alignment | 929 | 828 | 0 | 3 |
| Phase Alignment | 2,596 | 2,646 | 0 | 0 |
| CFR | 3,354 | 3,117 | 4 | 15 |
| DUC | 7,466 | 5,369 | 8 | 232 |
| DDC | 2,602 | 2,436 | 1 | 23 |
| Feedback Equalizer | 3,474 | 3804 | 0 | 38 |
| I/Q Compensation | 1,379 | 1,143 | 0 | 5 |
| SPI | 3,117 | 2556 | 0 | 0 |
| Total Used | 32,908 | 28,959 | 23 | 321 |
| Available | 407,600 | 203,800 | 1335 | 840 |
| Utilization | 8% | 14% | 1.7% | 38% |

for the modules of interest are listed. Two LUT units are employed for the DPD, and each consumes two block RAMs and one complex-number multiplier (3 multipliers). The predistortion only takes up a small portion of the whole used resources, while DUC accounts for most of the consumptions. The whole system only occupies 38% of the total DSP48A1s cores (multipliers).

## 5.4.1 LTE Signal Test

The nonlinear nature of the device under test (DUT) is apparent from the measured spectrum as depicted in Fig. 5.11. In this case, an LTE signal of 10 MHz carrier bandwidth was adopted as the input. It contained 600 subcarriers and employed 16-QAM for each subcarrier. The original PAPR was 11.2 dB (@ 0.0001% probability) and was reduced to 7.6 dB with a peak cancellation technique [68]. The backoff was set to 10 dB from the saturation point, and this results in average output power of around 41 dBm. The signal is steady, which means its average power is unchanged in the test procedure. It can be seen from the result presented in Fig. 5.11 that the ACLRs for lower and upper band are of -31.1 and -32.3 dB, respectively. The difference of the adjacent channel powers is 1.2 dB, as result of moderate memory effect. This result conforms to most practical cases where static nonlinearity is the dominate distortion. Consequently, the memory depth, viz. *M*, of the memory polynomial needs not to be high; and two LUT units were employed in the DPD. Furthermore, the first

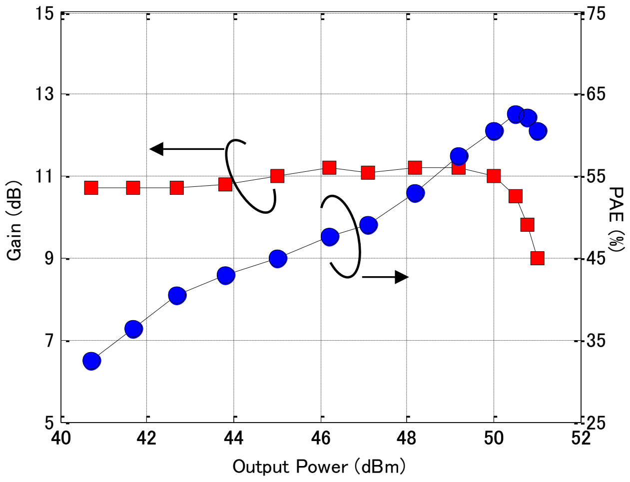


**Fig. 5.10 Measured gain and power-added efficiency of the Doherty amplifier.**

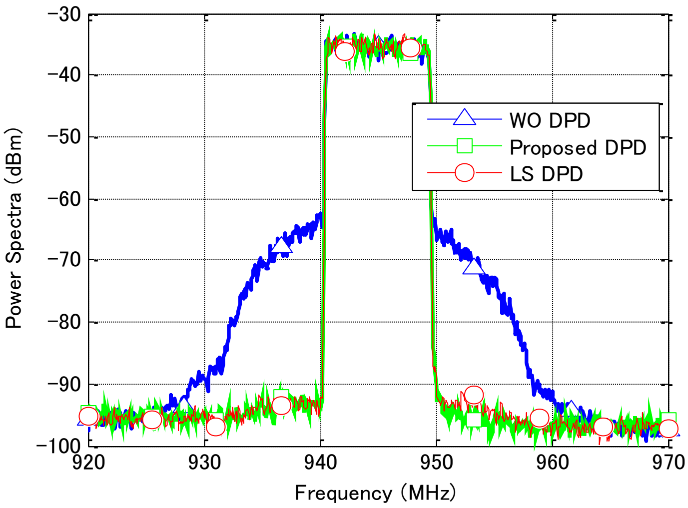


**Fig. 5.11 Power amplifier output spectrum plots for a 10 MHz LTE signal with and without DPD.**

LUT can be of larger size to accurately approximate the static nonlinearity, while smaller size can be assumed for the second LUT for the purpose of faster convergence.

Fig. 5.11 reports the test result of the power amplifier output spectrum applying the proposed DPD composing of two LUT units with sizes of 128 and 32. It is clear that the ACLR performance was improved by more than 28 dB and which reached -60.5 dB for both upper and lower bands.

The test result of applying a DPD based on least squares algorithm is also shown in Fig. 5.11. The least square fitting was performed in a micro processing unit (MPU) embedded in the FPGA. The input and output signals used for LS fitting were captured simultaneously with two 4096-size FIFOs implemented in the FPGA. The nonlinear order *N* and memory

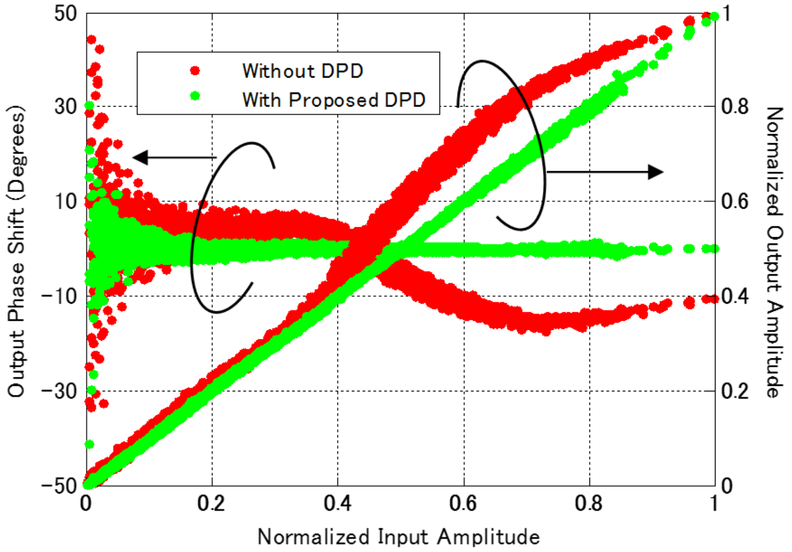


**Fig. 5.12 AM-AM and AM-PM characteristics of the amplifier with and without DPD.**

depth *M* of the memory polynomial were set to 6 and 2, respectively when extracting the coefficients. Then the 2 memoryless polynomials were mapped to 2 LUT units. The ACLR performances of the two DPDs are quite close, which implies that the proposed DPD can also achieve minimum quadratic error as the LS algorithm.

Fig. 5.12 shows the AM-AM and AM-PM characteristics of the PA with and without DPD. The input and output amplitudes were normalized by their respective maximum values. It can be seen from this figure that the proposed DPD can evidently suppress the nonlinear distortion consisting of both static nonlinearity and memory effect.

Fig. 5.13 shows the measured ACLR versus average output power, with and without the proposed DPD. As can be seen from the figure, the DPD achieves the best performance as the average output power approaches 40 dBm. The ACLR performance intensely deteriorates when the power exceeds 43 dBm. This is caused by the signal sweeping into the saturation region. On the other hand, the ACLR is also deteriorated in lower power region. This was caused by the SNR reduction as already discussed in Section 5.2.3.

Error vector magnitudes (EVMs) for different average output power were also measured and they are plotted in Fig. 5.14. Unlike the ACLR, EVM is a metric for the noise performance of the whole system. The CFR, which applies before the DPD, permanently degrades the EVM. In this test, the LTE signal after CFR has an EVM of 1.9 %, which represents the lower bound of the achievable EVM. In addition, the test bed, which includes modulator, demodulator, filters, etc., is also not ideal. With careful calibrations, such as the equalization discussed in Section 5.2.4., the EVM deterioration introduced by test bed can be kept within 0.5%. As can be seen from Fig. 5.14, the proposed DPD achieves EVM of around 2% in a wide range of power levels.

In order to view the convergence rate of our DPD, a RAM of a length of 16384 was employed. It started to store the instantaneous error signal as soon as the adaptation was

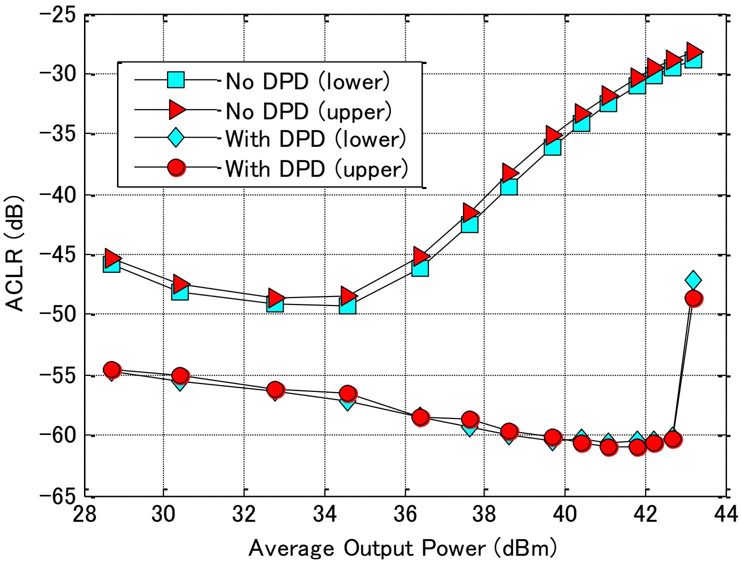


**Fig. 5.13** **Measured lower and upper ACLRs versus average output power with and without the proposed DPD.**

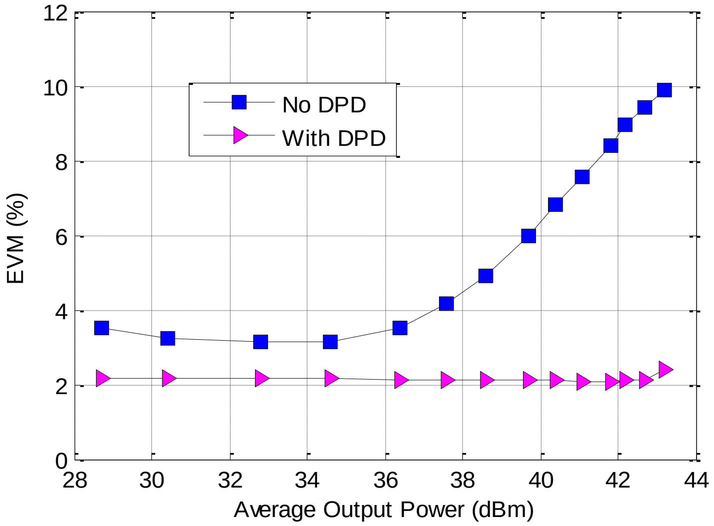


**Fig. 5.14** **Measured EVMs versus average output power with and without the proposed DPD.**

triggered. The error signal was decimated by a factor of 4, so that the recorded error is of a time span of $16384 \times 4/(92.16 \times 10^6) = 711\mu$s. The step size $\gamma$ for each adaptive LUT unit was set to 2, which is a good compromise for convergence rate and performance. The initial instantaneous error, i.e. learning curves for three different cases are depicted in Fig.5.15. The most rapid convergence, among the three cases, was obtained when the first and second LUT unit sizes were set to 64 and 32, respectively. With the first LUT of size 128, it gives slower convergence rate but lower error level. On the other hand, choosing smaller size for the

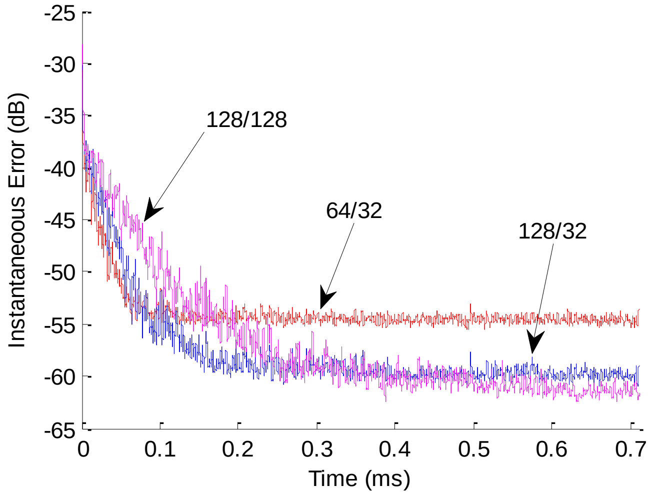


**Fig. 5.15 Initial instantaneous error of three cases: the sizes of the two LUT units are 64 and 32; the sizes of the two LUT units are 128 and 32; the sizes of the two LUT units are 128 and 128.**

second LUT can increase the convergence rate with negligible performance degradation.

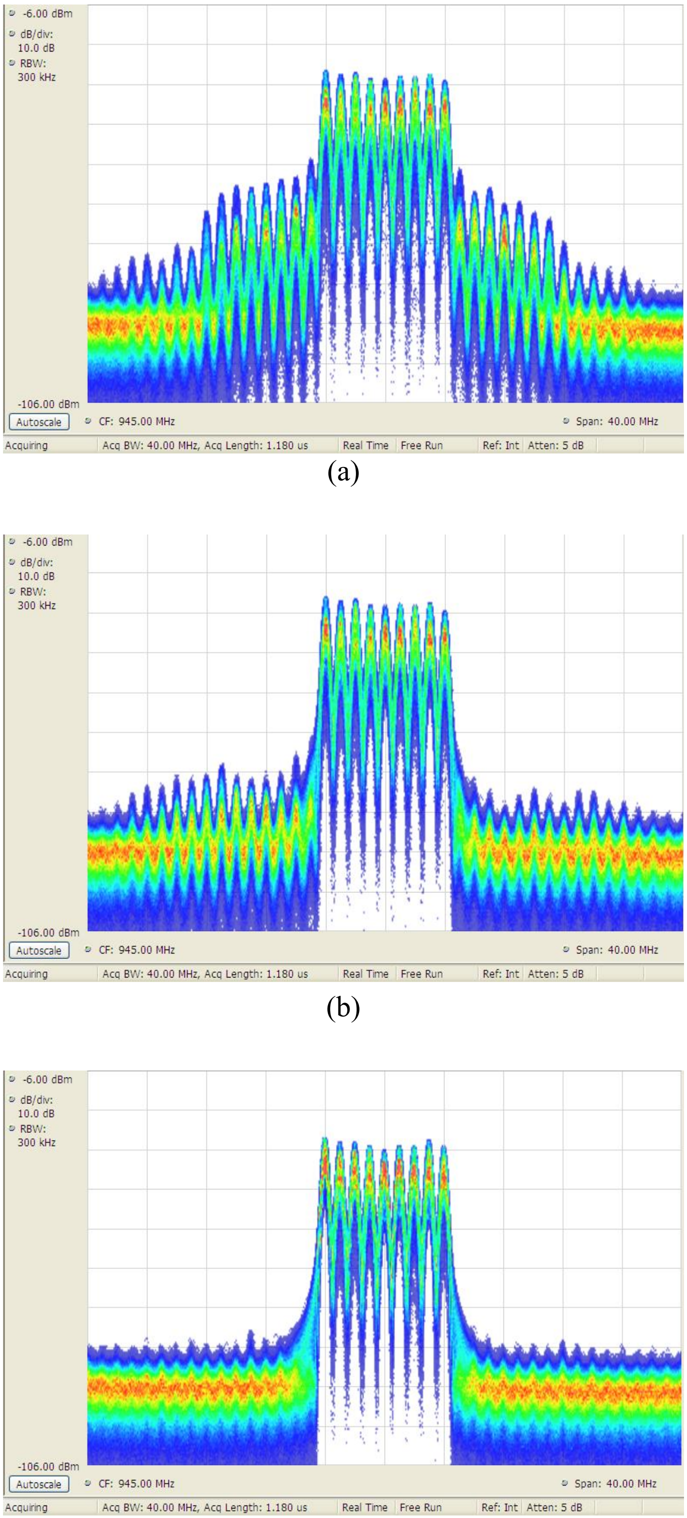
## 5.4.2 Multi-Channel GSM Signal Test

Because of the narrow bandwidth, the legacy GSM system allows more carriers (channels) being transmitted at the same time. This represents an even more difficult case for DPD than wideband systems, such as the LTE. The signal used in this test was a *dynamic* 9-carrier GSM signal with channel spacing of 1 MHz. The bandwidth of each channel is 200 kHz, and the carrier power varies from 20 dBm to 40 dBm.

The PA output spectrum, which was measured using a real-time spectrum analyzer, is plotted in Fig.5.16. Different colors in this figure denote different probabilities of the occurrence, e.g. red infers the highest probability. As shown in Fig.5.16(a), the varying channel power causes varying power distribution of inter-modulation distortion (IMD) components.

The test result for applying a DPD based on the least squares algorithm is shown in Fig.5.16(b). Since the least square algorithm is a one-step update and the signal characteristics had changed before the coefficients were updated to the DPD, there leaves obvious residual frequency contents as can be seen from the figure. It is justified by the test shown in Fig.5.16(c) that the performance of the proposed DPD acting with fast time varying signal is better than its counterpart employing LS algorithm. It can be seen that the out-of-band components were reduced by over 10 dB compared with Fig.5.16(b).

To emulate the signal encountered in the multi-channel transmitter with occasional carrier shutdown, the 9 channels were set to OFF or ON every 577$\mu$s, which is the span of one time slot for GSM, in a predefined pattern. This would produce a signal with fast varying in average power, probability distribution and PAPR. Fig. 5.17 demonstrates the procedure for generating the multi-channel GSM test signal with occasional channel carrier shutdown.



(a)

(b)

(c)

**Fig. 5.16 Spectrum plots of the PA output, where the signal is a 9-carrier GSM signal with channel power variation. From top to the bottom they are (a) without DPD; (b) with DPD employing LS algorithm; (c) with proposed DPD.**

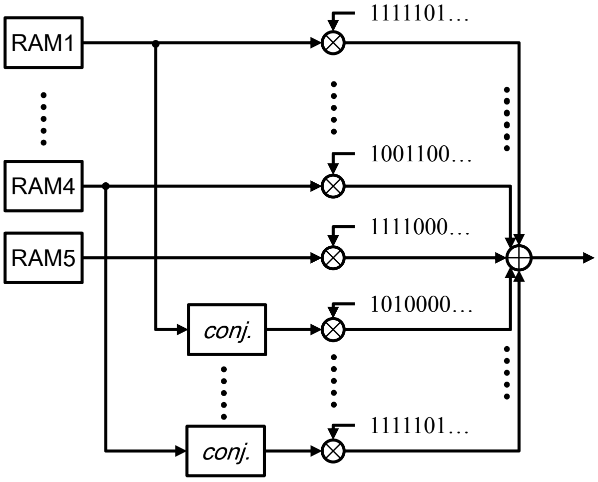
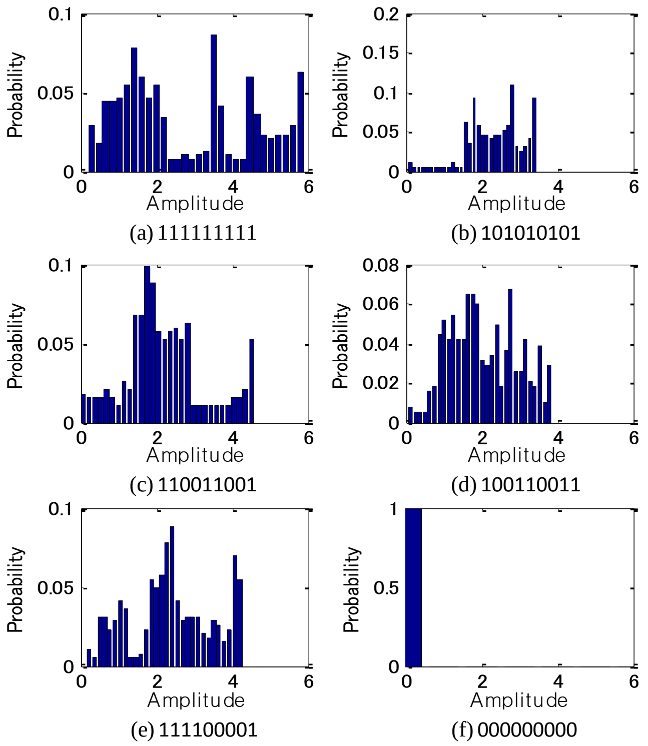


**Fig. 5.17 Demonstration of the procedure for producing the dynamic 9-carrier GSM signal.**

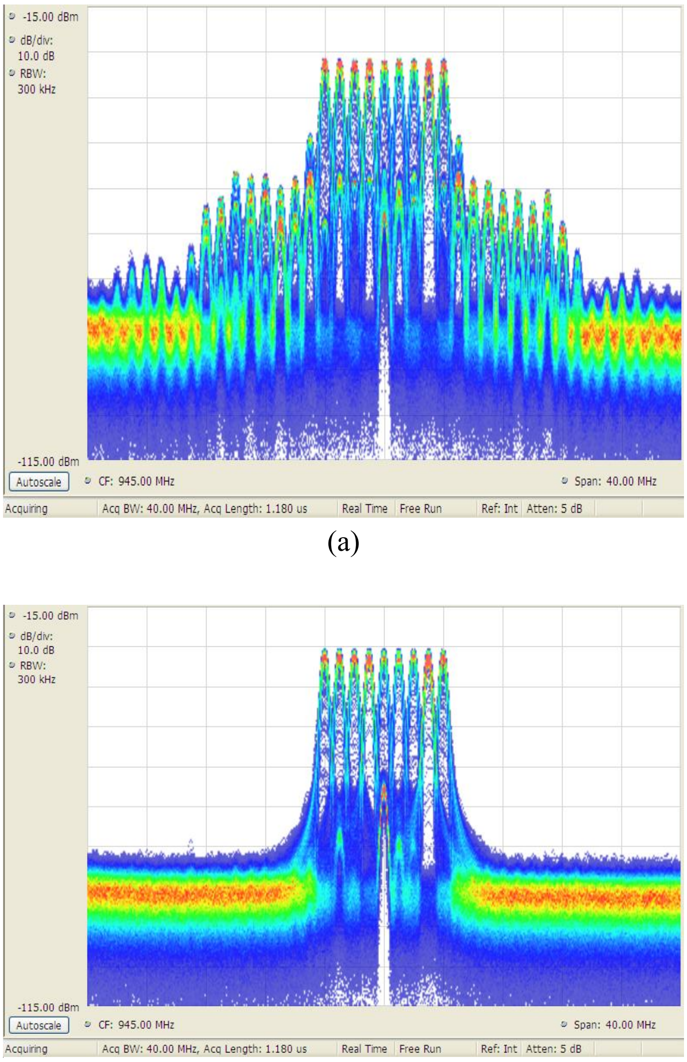


(a) 111111111
(b) 101010101
(c) 110011001
(d) 100110011
(e) 111100001
(f) 000000000

**Fig. 5.18 The probability histograms of the multi-carrier GSM signal amp- litudes which are calculated for each time slot. The bits shown below each figure denotes the ON/OFF states of the carriers.**

Five RAMs are used to store the complex signals at different frequencies, and the carriers in mirror frequencies are generated by taking the conjugation of the respective signals. The maximum PAPR of the multi-channel GSM signal is around 6.62dB. The ON and OFF states for each carrier are controlled by a list of predefined single bit stream, as shown in the figure.

With the control of channel carriers ON and OFF, the signal characteristics changes arduously at the onset of each time slot. The probability histograms of signal amplitudes for



(a)

(b)

**Fig. 5.19 Spectrum plots of the PA output, where the signal is a 9-carrier GSM signal with channel power shutdown. From top to the bottom they are (a) without DPD; (b) with proposed DPD.**

the first six time slots are shown in Fig.5.18. It can be noticed that in state (f), all channels are shut down to emulate the transmitter switching into sleeping mode. It is also worth to mention that the DPD based on LS algorithm is not workable for such signal, as it fails to track the arduous changes.

The spectrums both without and with the DPD are illustrate in Fig.5.19. It is clear that the channel power has even more violent variation compared with the case shown in Fig.5.15. The out-of-band emission can be suppressed by the proposed DPD, as can been seen from Fig. 5.19(b) that it almost reaches the noise floor. To view the instantaneous error of the proposed DPD in a larger time range, the error signal was decimated by 32 times in this test. As demonstrated in Fig.5.20, the DPD shows robust performance against the variation of the signal.

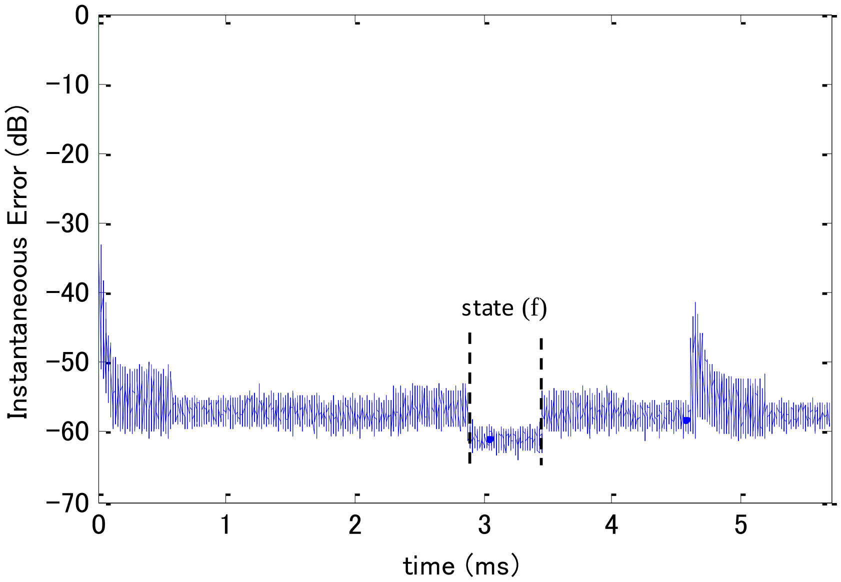


**Fig. 5.20 Initial instantaneous error when the test signal is the multi-channel GSM signal with channel shutdown.**

# Chapter 6.

# Ultra-Wideband Digital Predistortion Technology

With the ever increasing demands for higher data rate, wider bandwidth is required for improving the throughput. This trend, however, imposes design challenges for the digital predistortion (DPD) in many aspects. In order to sample the broadband power amplifier (PA) output signal, it requires the use of high-speed analog-to-digital converters (ADCs), which tend to be the most expensive components in a transmitter with DPD. The sampling speed of the ADC for conventional DPD has to be several times of the original signal bandwidth in order to cover the out-of-band intermodulation components caused by nonlinear PA. In this chapter, a novel technique, which allows the use of low-speed ADCs by introducing spectral extrapolation to the band-limited feedback signal, is presented. This allows efficient implementation of DPD for very wideband signals. Experimental results demonstrate that the bandwidth of the acquisition path can be even less than the bandwidth of original signal applying the proposed technique. In addition, satisfactory linearization performance has been achieved employing wideband signals up to 320 MHz bandwidth.

## 6.1 DPD WITH LOWER SPEED ADC

The explosive growth of the demand on higher data rates is calling for wider utilization bandwidth to support higher throughput. For example, carrier aggregation (CA) specified to aggregate multiple long term evolution (LTE) carrier components (CCs), which is a key technology known for LTE-Advanced, allows maximum of 5 CCs to be simultaneously supported in a continuous band. This represents a contiguous spectrum of 100 MHz bandwidth [119]. This trend has a large impact on DPD design in many aspects, not the least of which is the wide band signal to be processed. Generally, the DPD system contains ADCs to sample the PA output and feed it back to the DPD, see Fig.6.1 in which the transmission path and acquisition path are both based on direct conversion structure. The acquisition path for conventional DPD, which contains a down converter, anti-aliasing filters and ADCs, should cover a span several times of the original signal bandwidth, inasmuch as the nonlinearity of the PA expands the spectrum much wider than the original signal. For an acquisition path that is based on direct conversion structure, for instance, the sampling rate of the ADCs

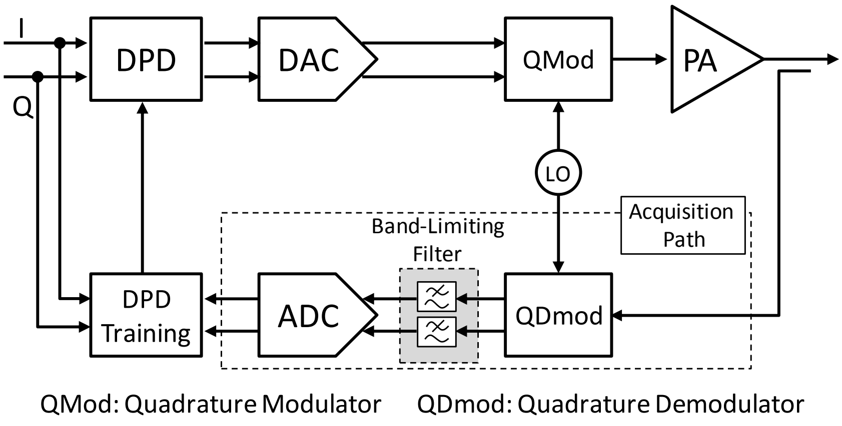


**Fig. 6.1 The diagram of DPD considering the effect of band-limiting filters in the acquisition path.**

should be at least 500 MSa/s if the aforementioned LTE-Advanced signal is applied. However, high-speed ADC tends to be one of the most expensive and power-hungry components for a commercial transmitter. Consequently developing a DPD using low-speed ADCs, while maintaining reasonable good performance, would be rather cost-effective and quite meaningful from the practical point of view.

There are several known methods to perform predistortion with low-speed ADCs. One is the so-called scalar DPD, in which only the out-of-band power is monitored, while the DPD update is aiming to minimize the out-of-band power [120]. The drawbacks of this method are slow convergence rate and incapability for memory effect. Another method [121] is based on Zhu’s generalized sampling theory, which justifies that sampling the signal with the input Nyquist rate suffices to model the memoryless nonlinear system. The basic idea of this method is to undersample the PA output and carry out proper interpolations to obtain high-rate model. However, memory effect, which can be interpreted as frequency domain unflatness, cannot be well approximated with aliased spectra. Another method applies band-limiting filters to the model and constrains the error in the bandwidth of the filters [122]. It requires the use of a bulky filter behind the PA, as the error is only minimized in the prescribed frequency span. The authors in [123] found that direct learning DPD shows better performance than its indirect learning counterpart when band-limited feedback signal is employed.

In this chapter, we considers a predistorter with band-limited acquisition path, see Fig.6.1. As the spectra outside the acquisition band are filtered out by the band-limiting filters, the demands on ADC speed can be much lower for obtaining un-aliased signal. To recover the spectra outside the acquisition band, spectral extrapolation is introduced. A dumped Gauss Newton algorithm is proposed to adaptively find the optimum DPD parameters. With the proposed method, the bandwidth of the acquisition path can be even less than the bandwidth of original signal as shown by the experimental results. It also shows that with an acquisition path of 320 MHz bandwidth, the proposed DPD has achieved good linearization performance when a signal with a bandwidth of over 320 MHz was applied. To the best of our knowledge, there exists no other DPD technique in literatures has the ability to linearize the PA when such a wideband signal is applied. As a result, the proposed DPD can be seen as a promising technique for PA linearization for ultra wideband systems.

## 6.2 FEEDBACK SIGNAL RESTORATION USING SPECTRAL EXTRAPOLATION

### 6.2.1 Problem Formulation and Mathematical Notations

In order to cast the problem at hand into more concrete mathematical terms, it is beneficial to give it an algebraic structure. The involved signals can be expressed by $N$-tuple complex vectors $\mathbf{x}, \mathbf{y}, \mathbf{y}_B \in \mathbb{C}^N$, which represent the original input signal, distorted PA output signal and band-limited feedback signal, respectively. $\mathbf{y}_B$ is said band-limited if its discrete Fourier transform (DFT) $\mathbf{Y}_B \in \mathbb{C}^N$ satisfies

$$\mathbf{Y}_B(m) = 0 \qquad |m| > \frac{M}{2}, \tag{6.1}$$

where $M$=$N$B$_l$ is the number of in-band [3]spectral elements ($M$ is assumed to be an integer), if B$_l$ is the normalized bandwidth of the band-limiting filter. DFT can be accomplished by left-multiplying the prescribed signal vector with a transformation matrix $\mathbf{W}_N$, which has the form of

$$\mathbf{W}_N(p,q) = \left( \frac{\omega^{pq}}{\sqrt{N}} \right)_{p,q=-\frac{N}{2}+1,\cdots,\frac{N}{2}}, \tag{6.2}$$

in which

$$\omega = e^{-\frac{2\pi j}{N}} \tag{6.3}$$

is a primitive $N$-th root of unity [124]. An inverse DFT (IDFT) transformation matrix $\tilde{\mathbf{W}}_N$ can be defined in a similar way. We further define three matrices:

$$\begin{aligned} \mathbf{H}_i &= [\mathbf{0} \quad \mathbf{I}_M \quad \mathbf{0}] \in \mathbb{R}^{M \times N} \\ \mathbf{H}_l &= [\mathbf{I}_L \quad \mathbf{0}] \in \mathbb{R}^{L \times N} \\ \mathbf{H}_u &= [\mathbf{0} \quad \mathbf{I}_L] \in \mathbb{R}^{L \times N}, \end{aligned} \tag{6.4}$$

---

[3] The term *in-band* is used to denote the spectral signals inside the acquisition band and conversely *out-band* to denote the signals outside the corresponding bandwidth.

where $L$=($N$-$M$)/2 is the number of out-band elements for lower and upper bands. $\mathbf{I}_L$ and $\mathbf{I}_M$ are $L\times L$ and $M\times M$ unit matrices and **0** represents matrix of zeros. The out-band signals for lower and upper bands in time domain can be obtained with such operations:

$$\begin{aligned}\mathbf{y}_l &= \tilde{\mathbf{W}}_L\mathbf{H}_l\mathbf{W}_N\mathbf{y} = \mathbf{P}_l\mathbf{y}\\ \mathbf{y}_u &= \tilde{\mathbf{W}}_L\mathbf{H}_u\mathbf{W}_N\mathbf{y} = \mathbf{P}_u\mathbf{y}.\end{aligned} \tag{6.5}$$

$\mathbf{P}_l$ and $\mathbf{P}_u$ are projection operators:

$$\begin{aligned}\mathbf{P}_l &: \Omega_N \to \Omega_l\\ \mathbf{P}_u &: \Omega_N \to \Omega_u,\end{aligned} \tag{6.6}$$

where $\Omega_N \subseteq \mathbb{C}^N$ and $\Omega_l, \Omega_u \subseteq \mathbb{C}^L$ are subspaces. A projection operator $\mathbf{P}_i : \Omega_N \to \Omega_i$ ($\Omega_i \subseteq \mathbb{C}^M$) can be defined in a similar way:

$$\mathbf{P}_i = \tilde{\mathbf{W}}_M\mathbf{H}_i\mathbf{W}_N, \tag{6.7}$$

and

$$\hat{\mathbf{y}} = \mathbf{P}_i\mathbf{y} \tag{6.8}$$

represents the corresponding in-band signal in time domain. If the ADC is assumed to be of the sampling rate coherent with the bandwidth of the filter, $\hat{\mathbf{y}}$ would also represent the ADC output. It is easy to see that

$$\hat{\mathbf{y}} = \mathbf{P}_i\mathbf{y}_B \tag{6.9}$$

if the band-limiting filter is ideal, i.e. the filter eliminates all out-band frequencies while passing those in-band unchanged.

To recover **y** from the incomplete observation $\hat{\mathbf{y}}$ is a signal extrapolation problem, in other words, is to obtain a projector $\Xi : \Omega_i \to \Omega_N$ such that

$$\mathbf{y}' = \Xi\hat{\mathbf{y}} \tag{6.10}$$

approaches or equals to **y**. It is obvious that the following equation should be met:

$$\Xi\mathbf{P}_i \cong \mathbf{I}_N, \tag{6.11}$$

where $\mathbf{I}_N$ is an $N\times N$ unit matrix.

### 6.2.2 One-Step Spectral Extrapolation Method

Complete feedback signal should be restored in order to extract the predistorter coefficients correctly. This is to extrapolate the incomplete feedback signal spectrum outside the acquisition bandwidth. General signal extrapolation refers to the problem of finding an estimate of a signal outside its observation interval, which can be formulated as an inverse problem in Hilbert space [125][126].

Consider that the power amplifier is a continuous weak nonlinear function with memory, $x$ and $y$ can be related with the well-established Volterra series, which in the discrete time domain, is given by

$$y(n) = \sum_{p=1,3,5\cdots}^{\infty} \sum_{i_p=0}^{\infty} \left[ a_n(i_p) x(n-i_1) \times \prod_{m=1}^{\infty} x(n-i_{2m}) x^*(n-i_{2m+1}) \right] \tag{6.12}$$

where $a_n(i_p)$ is the $p$-th order Volterra kernel and

$$k_{i_p}\left[x(n)\right] = x(n-i_1) \prod_{m=1}^{\infty} x(n-i_{2m}) x^*(n-i_{2m+1}) \tag{6.13}$$

is the Volterra operator. Consider a Volterra series with finite orders, say, $P$ operators, (6.13) can be expressed as

$$\mathbf{y} = k(\mathbf{x})\mathbf{a}, \tag{6.14}$$

where $k(\mathbf{x})=[k_1(\mathbf{x}), k_2(\mathbf{x}),\ldots, k_P(\mathbf{x})]$ is the function space, and the vectors $k_1(\mathbf{x}), k_2(\mathbf{x}),\ldots, k_P(\mathbf{x})$ are the linear regressors. In algebraic terms, alternatively, we have

$$\mathbf{y} \in \operatorname{ran}\left(k(\mathbf{x})\right), \tag{6.15}$$

which denotes that $\mathbf{y}$ is in the range of $k(\mathbf{x})$. For signal extrapolation problems, *a priori* information describing the distribution of the signal energy in time/frequency domain is needed. The energy-concentrated extrapolation incorporates such knowledge of either time-domain or frequency-domain by minimizing the energy in the constrained span other than the total span [126]. Here, the band-limiting filter is known to reflect this concentration, and it is easy to see that

$$\hat{\mathbf{y}} \in \operatorname{ran}(\mathbf{P}_i k(\mathbf{x})). \tag{6.16}$$

Therefore, the problem at hand is to estimate the sequence in the null space of $\mathbf{P}_i k(\mathbf{x})$, and infinite solutions may exist. In order to choose a single extrapolation from this infinite set of solutions, additional assumptions are added concerning the nature of the observed signal. One popular minimum-energy solution for $\boldsymbol{\Xi}$, which has been derived in a variety of fashions, minimizes the sum of squares of sequence values

$$\left\| \mathbf{y}' - \boldsymbol{\Xi}\hat{\mathbf{y}} \right\|_2^2, \tag{6.17}$$

can be obtained with the one-step method [14][15]:

$$\boldsymbol{\Xi} = k(\mathbf{x})\left(k(\mathbf{x})^H \mathbf{P}_i^H \mathbf{P}_i k(\mathbf{x})\right)^{-1} k(\mathbf{x})^H \mathbf{P}_i^H. \tag{6.18}$$

As (6.18) extrapolates the frequency representation of the complete response, both static nonlinearity and dynamic nonlinearity (memory effect) can be retrieved. Thus, the extrapolation can potentially provide significant advantages for PAs with memory effect, compared with existing techniques, such as [120] and [121], which lack of the information outside the acquisition bandwidth.

### 6.2.3 Compensation of the Non-Ideal Band-Limiting Filter

The discussion before has assumed ideal band-limiting filter, which, especially for wideband applications, is difficult to be ensured. Under the circumstance of non-ideal filter, the actual feedback signal can be expressed by

$$\mathbf{y}_f = \mathbf{H}\mathbf{y} \in \mathbb{C}^N, \tag{6.19}$$

where **H** is a linear operator and it has the Toeplitz form of

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(1) & \mathbf{h}(2) & \cdots & \mathbf{h}(N) \\ \mathbf{h}(N) & \mathbf{h}(1) & \cdots & \mathbf{h}(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}(2) & \mathbf{h}(3) & \cdots & \mathbf{h}(1) \end{bmatrix}, \tag{6.20}$$

and **h** is the impulse response (after zero-padding if necessary) of the actual band-limiting filter.

With the effect of **H**, the actual ADC output becomes

$$\hat{\mathbf{y}} = \mathbf{P}_i\mathbf{y}_f = \mathbf{P}_i\mathbf{H}\mathbf{y}. \tag{6.21}$$

It hinders the extrapolation from recovering **y**, as $\mathbf{P}_i\mathbf{H}$ is not a projection. This is obvious because it is not idempotent; more specifically, the non-ideal filter produces dissimilar signal as doing it twice. Therefore, the effect of **H** should be eliminated and this can be accomplished by using an extra compensation filter.

Here we assume the impulse response of the compensation filter to be $\mathbf{h}_c$ and its Toeplitz form to be denoted as $\mathbf{H}_c$. Instead of compensating the actual band-limiting filter in the whole frequency span, $\mathbf{h}_c$ only needs consider the in-band part. i.e.

$$\mathbf{H}\mathbf{h}_c \overset{set}{=} \tilde{\mathbf{W}}_N\mathbf{d}, \tag{6.22}$$

where

$$\mathbf{d} = \underbrace{[0 \ \cdots \ 0 \quad \overbrace{1 \ \cdots \ 1}^{M} \quad 0 \ \cdots \ 0]}_{N}{}^T \tag{6.23}$$

represents the ideal frequency response of the band-limiting filter. Equation (6.22) implies that connecting the two filters in tandem produces an impulse response for ideal band-limiting filter, and the least square solution for $\mathbf{h}_c$ can be given as

$$\mathbf{h}_c = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\tilde{\mathbf{W}}_N\mathbf{d}. \tag{6.24}$$

Applying the compensation filter to the actual feedback signal will produce

$$\mathbf{H}_c\mathbf{y}_f = \mathbf{H}_c\mathbf{H}\mathbf{y} \approx \mathbf{y}_B, \tag{6.25}$$

so that the response of the acquisition path can equal to an almost ideal band-limiting filter. Therefore, the band-limiting filter is assumed ideal in the following discussion.

## 6.3 PREDISTORTION USING SPECTRAL EXTRAPOLATION OF BAND-LIMITED FEEDBACK SIGNAL

With spectral extrapolation, the tools have been available to build a predistorter applying low-rate ADCs. As the DPD is the pre-inverse of the PA, its coefficients estimation becomes a nonlinear problem. The indirect learning method, which casts the nonlinear problem into linear one, is to estimate the post-inverse instead, and copy the post-inverse to the DPD. As the post-inverse is linear in respective to its coefficients, standard linear estimation methods, such as least squares (LS), can be applied to extract the coefficients. However, for indirect learning there also exists some issues need to be carefully considered. One of which is the power alignment problem as noticed by the author in [109]. The coefficients obtained by LS are not the optimum solution, as the DPD is a nonlinear function, which gives rise to increase of peak-to-average power ratio (PAPR) after the DPD coefficients are updated.

An alternative scheme is the direct learning, which as its name implies, directly estimates the pre-inverse of the PA. It has lower complexity but represents a nonlinear optimization problem. In this section, a technique that combines spectral extrapolation and direct identification of the DPD will be presented.

To start with, we consider the output of the PA with DPD: $\mathbf{y}=f(\mathbf{v})$, where $f(\,)$ is the function of the PA and $\mathbf{v}\in\mathbb{C}^N$ is the DPD output. If a $p$-th order Volterra series is applied as the DPD model, the PA output is given as

$$\mathbf{y}=f\left(k(\mathbf{x})\boldsymbol{\beta}\right), \tag{6.26}$$

where $\boldsymbol{\beta}\in\mathbb{C}^P$ contains the DPD coefficients. Normally, the objective of the DPD is to minimize the quadratic Euclidean norm $\|\mathbf{x}-\mathbf{y}\|_2^2$, whereas in our case, only incomplete version of $\mathbf{y}$ is available. Therefore, we define the following error function instead

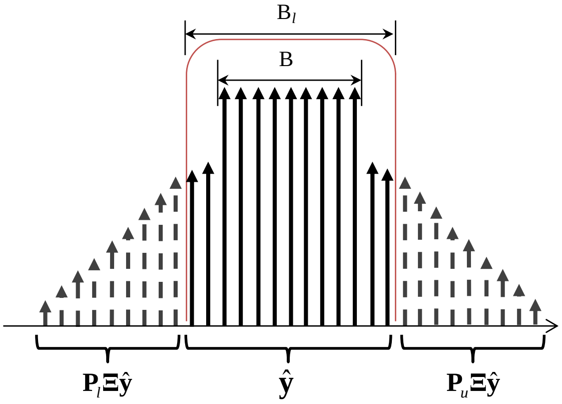
$$R(\boldsymbol{\beta})=\mathbf{D}_w\begin{bmatrix}\xi_l\\ \xi_i\\ \xi_u\end{bmatrix}=\mathbf{D}_w\xi, \tag{6.27}$$

where

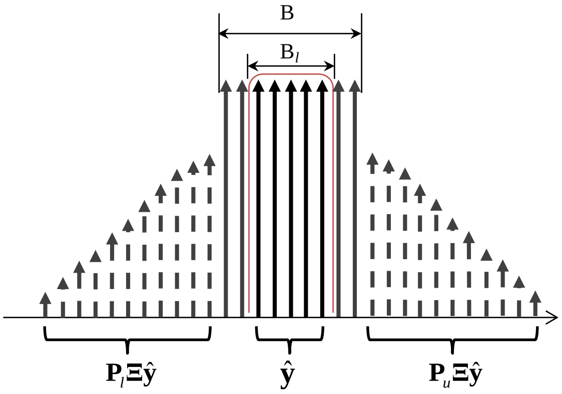
$$\begin{aligned}\xi_i&=\mathbf{P}_i\mathbf{x}-\hat{\mathbf{y}}\\ \xi_l&=-\mathbf{P}_l\Xi\hat{\mathbf{y}}\\ \xi_u&=-\mathbf{P}_u\Xi\hat{\mathbf{y}}\end{aligned} \tag{6.28}$$

denote the in-band error and out-band errors for lower and upper bands, respectively. $\xi_i$, $\xi_l$ and $\xi_u$ should not be given equal priority, as $\xi_l$ and $\xi_u$, which are obtained by extrapolation, normally have much lower power and less reliability. Thus, a matrix $\mathbf{D}_w\in\mathbb{R}^{N\times N}$ is applied in (6.21) to cast different weights on the error function, i.e. to give $\xi_i$ higher priority.

Notice that if the acquisition bandwidth is smaller than the original signal bandwidth,



(a) The original signal bandwidth is less than the acquisition bandwidth



(b) The original signal bandwidth is larger than the acquisition bandwidth

**Fig. 6.2 Demonstration of how the criteria function is defined.**

the projectors $\mathbf{P}_l$ and $\mathbf{P}_u$ in the above function need to be redefined. To facilitate understanding, the two cases are illustrated in Fig.6.2. When the original signal bandwidth is covered by the band-limiting filter, see Fig.6.2(a), the projectors conform to the one we have defined in last section. Whereas for the second case shown in Fig.6.2(b), in which the acquisition bandwidth is less than the original signal bandwidth, $\mathbf{P}_l$ and $\mathbf{P}_u$ should only select the spectral contents outside the signal bandwidth. Therefore, the $\mathbf{P}_l$ and $\mathbf{P}_u$ can be modified by redefining $\mathbf{H}_l$ and $\mathbf{H}_u$:

$$\begin{aligned}\mathbf{H}_i &= [\mathbf{0} \quad \mathbf{I}_M \quad \mathbf{0}] \in \mathbb{R}^{M\times N}\\ \mathbf{H}_l &= [\mathbf{I}_S \quad \mathbf{0}] \in \mathbb{R}^{S\times N}\\ \mathbf{H}_u &= [\mathbf{0} \quad \mathbf{I}_S] \in \mathbb{R}^{S\times N},\end{aligned} \tag{6.29}$$

where S=(1-max[B, B$_l$])$N/2$. B and B$_l$ are the normalized signal bandwidth and acquisition bandwidth, respectively. On the other hand, $\mathbf{D}_w$ can be obtained using the following equation

$$\mathbf{D}_w = \hat{\mathbf{W}}_N \begin{bmatrix} \mathbf{D}_l & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{D}_i & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{D}_u \end{bmatrix}_{N\times N} \mathbf{W}_N, \tag{6.30}$$

where diagonal matrices $\mathbf{D}_l, \mathbf{D}_u \in \mathbb{R}^{S\times S}$ and $\mathbf{D}_i \in \mathbb{R}^{M\times M}$ are responsible for the out-band and in-band contents, respectively.

The nonlinear least square (NLS) estimator applies a criteria function obtaining $\boldsymbol{\beta}$ that minimizes

$$\begin{aligned} J(\boldsymbol{\beta}) &= R^H(\boldsymbol{\beta})R(\boldsymbol{\beta}) \\ &= \|\mathbf{D}_w\boldsymbol{\xi}\|_2^2 \end{aligned} \tag{6.31}$$

is generally an iterative procedure:

$$\boldsymbol{\beta}^{(q+1)} = \boldsymbol{\beta}^{(q)} + \mathbf{s}^{(q)}\mathbf{d}^{(q)}, \tag{6.32}$$

so that the ($q$+1)-th iterated value $\boldsymbol{\beta}^{(q+1)}$ is obtained from the previous iteration value $\boldsymbol{\beta}^{(q)}$, by adjusting the amount $\mathbf{s}^{(q)}\mathbf{d}^{(q)}$, where $\mathbf{d}^{(q)}$ characterizes the direction of change in parameter space and $\mathbf{s}^{(q)}$ controls the amount of change. There are several different algorithms for NLS problem, with respective abilities to achieve local or global minimization [127], which require the computation of the first order difference (Jacobian) or/and second order difference (Hessian) of the criteria function. For example, the damped Gauss-Newton algorithm with the form of

$$\boldsymbol{\beta}^{(q+1)} = \boldsymbol{\beta}^{(q)} + \frac{\gamma}{2}\left(R'(\boldsymbol{\beta})^H R'(\boldsymbol{\beta})\right)^{-1} J'(\boldsymbol{\beta}) \tag{6.33}$$

has a descent direction. The parameter $\gamma$ controls the step size and a minimum can be found by one time of iteration if it is set to 1. $R'(\boldsymbol{\beta})$ and $J'(\boldsymbol{\beta})$ are the gradients of $R(\boldsymbol{\beta})$ and $J(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. For predistortion problems, nevertheless, the function of the PA $f(\mathbf{v})$ is unknown, which in turn perplexes the application of standard NLS algorithms, as it demands the Taylor expansion of $f(\mathbf{v})$ for calculation of the Jacobian. Fortunately, the typical PA is a weak nonlinear system whose 3rd and higher order Taylor series has much lower amplitude than the linear component. Consequently, such an approximation can be made:

$$\hat{\mathbf{y}} = \mathbf{P}_i f(\mathbf{v}) \approx \mathbf{P}_i k(\mathbf{x})\boldsymbol{\beta}. \tag{6.34}$$

By combining (6.27), (6.28) and (6.34) there yields

$$R(\boldsymbol{\beta}) \approx \mathbf{D}_w \begin{bmatrix} -\mathbf{P}_l \boldsymbol{\Xi} \mathbf{P}_i k(\mathbf{x})\boldsymbol{\beta} \\ \mathbf{P}_i\mathbf{x} - \mathbf{P}_i k(\mathbf{x})\boldsymbol{\beta} \\ -\mathbf{P}_u \boldsymbol{\Xi} \mathbf{P}_i k(\mathbf{x})\boldsymbol{\beta} \end{bmatrix}. \tag{6.35}$$

Note that $\boldsymbol{\Xi}$ is the pseudo-inverse of $\mathbf{P}_i$ and by recall (6.11), we have

$$R(\boldsymbol{\beta}) \approx \mathbf{D}_w \begin{bmatrix} -\mathbf{P}_l k(\mathbf{x})\boldsymbol{\beta} \\ \mathbf{P}_i\mathbf{x} - \mathbf{P}_i k(\mathbf{x})\boldsymbol{\beta} \\ -\mathbf{P}_u k(\mathbf{x})\boldsymbol{\beta} \end{bmatrix} \tag{6.36}$$

and

$$\begin{aligned} R'(\boldsymbol{\beta}) &= \frac{\partial R(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ &\approx \mathbf{D}_w \begin{bmatrix} -\mathbf{P}_l \\ -\mathbf{P}_i \\ -\mathbf{P}_u \end{bmatrix} k(\mathbf{x}) \\ &= \mathbf{D}_w \mathbf{P} k(\mathbf{x}), \end{aligned} \tag{6.37}$$

where

$$\mathbf{P} = -\begin{bmatrix} \mathbf{P}_l \\ \mathbf{P}_i \\ \mathbf{P}_u \end{bmatrix}. \tag{6.38}$$

Therefore

$$\begin{aligned} J'(\boldsymbol{\beta}) &= \frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ &\approx 2k(\mathbf{x})^H \mathbf{P}^H \mathbf{D}_w^H R(\boldsymbol{\beta}) \\ &= 2k(\mathbf{x})^H \mathbf{P}^H \mathbf{D}_w^H \mathbf{D}_w \boldsymbol{\xi}. \end{aligned} \tag{6.39}$$

By substituting (6.37) and (6.39) into (6.33) it yields

$$\begin{aligned} \boldsymbol{\beta}^{(q+1)} = \boldsymbol{\beta}^{(q)} + \gamma \left( k(\mathbf{x})^H \mathbf{P}^H \mathbf{D}_w^H \mathbf{D}_w \mathbf{P} k(\mathbf{x}) \right)^{-1} \\ k(\mathbf{x})^H \mathbf{P}^H \mathbf{D}_w^H \mathbf{D}_w \boldsymbol{\xi}. \end{aligned} \tag{6.40}$$

The spectral extrapolation procedure, as consequence, is embedded into the damped Gauss-Newton iteration.

To enable real-time operation, running signal blocks are used to replace the signal vectors:

$$\begin{aligned} \boldsymbol{\beta}^{(q+1)} = \boldsymbol{\beta}^{(q)} + \gamma \left( k(\mathbf{x}^{(q)})^H \mathbf{P}^H \mathbf{D}_w^H \mathbf{D}_w \mathbf{P} k(\mathbf{x}^{(q)}) \right)^{-1} \\ k(\mathbf{x}^{(q)})^H \mathbf{P}^H \mathbf{D}_w^H \mathbf{D}_w \boldsymbol{\xi}^{(q)}, \end{aligned} \tag{6.41}$$

where

$$\boldsymbol{\xi}^{(q)} = \begin{bmatrix} -\mathbf{P}_l \boldsymbol{\Xi}^{(q)} \hat{\mathbf{y}}^{(q)} \\ \mathbf{P}_i \mathbf{x}^{(q)} - \hat{\mathbf{y}}^{(q)} \\ -\mathbf{P}_u \boldsymbol{\Xi}^{(q)} \hat{\mathbf{y}}^{(q)} \end{bmatrix} \tag{6.42}$$

and

$$\boldsymbol{\Xi}^{(q)} = k(\mathbf{x}^{(q)}) \left( k(\mathbf{x}^{(q)})^H \mathbf{P}_i^H \mathbf{P}_i \, k(\mathbf{x}^{(q)}) \right)^{-1} k(\mathbf{x}^{(q)})^H \mathbf{P}_i^H \tag{6.43}$$

are involved matrices for $q$-th iteration. The signal vectors have the form of

$$\mathbf{x}^{(q)} = \begin{bmatrix} \mathbf{x}\left((q-1)N+1\right) \\ \mathbf{x}\left((q-1)N+2\right) \\ \vdots \\ \mathbf{x}\left(qN\right) \end{bmatrix} \tag{6.44}$$

and $\hat{\mathbf{y}}^{(q)}$ can be defined in a similar manner. The iteration stops as the criteria function is below a predefined value.

With block-by-block operation, damped Gauss-Newton maximizes the efficiency of the system by distributing the fixed process overhead across many samples. In addition, damped Gauss-Newton absorbs the merits of LS, such as high robustness, and enables adaptive update of the DPD. Also, the length of vectors involved can be much smaller than the ones in LS, hence the required memory space and computation load are much lower.

## 6.4 EXPERIMENTAL RESULTS

### 6.4.1 Description of the Measurement

Experiments were taken place to evaluate the suitability of the proposed method. The measurement applied a power amplifier, which is a general purposed wideband PA HMC-C074 from Hittite corporation. The test bed (see Fig.6.3) is comprised of an arbitrary signal generator (AWG) M8190A as the transmitter and a signal analyzer PXA N9030 as the receiver. The M8190A, which has a sampling rate of maximum 12 GSa/s and an analog bandwidth of 5 GHz, can directly output the digitalized RF (radio frequency) signal. The baseband signal is transferred to RF by digital up conversion (DUC) that is implemented in the MATLAB. The N9030, which contains the acquisition path, supports a sampling rate up to 200 MSa/s and its intermediate frequency (IF) filter bandwidth is 160 MHz. The two instruments are controlled with the MATLAB instrument toolbox through PCIe and GBIP ports, respectively. The spectrum is displayed on a spectrum analyzer, which is an FSV from Rohde & Schwarz. The step size $\gamma$ is set to 0.5 and the length of block $N$ is set to 512 for the iterative algorithm. The baseband memory polynomial model that has defined in (3.7)

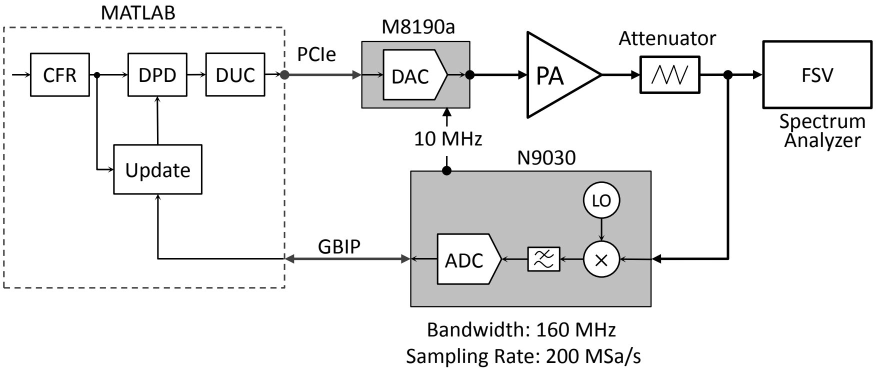


**Fig. 6.3 Outline of the measurement setup.**

$$v(n) = \sum_{m=0}^{2} \sum_{p=0}^{3} |x(n-m)|^{2p} x(n-m), \tag{6.45}$$

which is a pruned version for Volterra series, was applied as the DPD model. It is worth to mention that the *even order terms*, which are claimed beneficial in [85], should not be included in this circumstance. This is because the *even order terms* are actually strong nonlinear functions (consider the complex number absolution), which are unsuitable for the band-limited signal extrapolation as their outputs generally have very wide spectrum.

The LTE signal used in our test employs OFDM, which, nevertheless, has a side-effect of high peak-to-average power ratio (PAPR), as already discussed in Section 2.1.3. The increased PAPR will mandate the signal to work with larger backoff after the DPD taking effect. As a result, crest factor reduction (CFR) becomes mandate to facilitate the DPD when signal of high PAPR is coped.

The occurrence of high peaks is a statistic issue, which for LTE signals, is infrequent enough that it is acceptable to be removed by another block. This chapter applies peak cancellation, which provides more tractable PAPR reduction as mentioned in 2.1.3. Compromise should be made to balance the performance of the whole system, as the adjacent channel leakage ratio (ACLR) and error vector magnitude (EVM) may be degraded by CFR.

## 6.4.2 Performance with LTE Signal

The test signal is an LTE signal, which is a 64-QAM OFDM signal with 1200 occupied subcarriers within 18 MHz occupied bandwidth and 20 MHz channel bandwidth. As the test setup has an acquisition bandwidth of 160 MHz, a digital filter was added to emulate the effect of the band-limiting filter. The original LTE signal has a PAPR of around 11 dB (@ 0.00001% probability) which was strictly constrained below 7.5 dB with the peak cancellation. A set of tests was performed to evaluate the performance of the propose method under different band-limiting filters. The average output power was set to 20 dBm, which

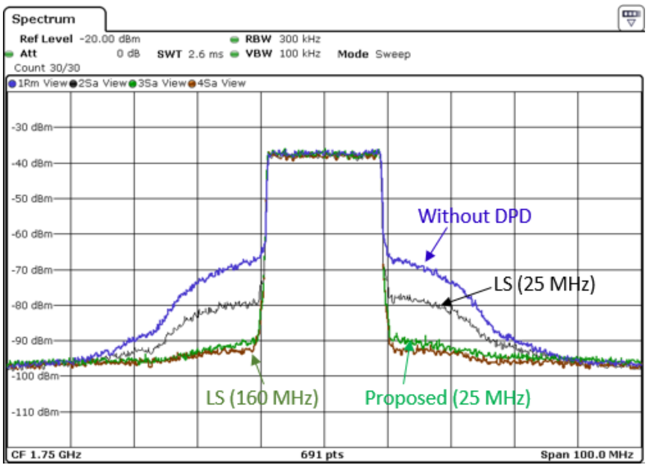


**Fig. 6.4 Comparison of the PA output spectra obtained from the proposed method and LS.**

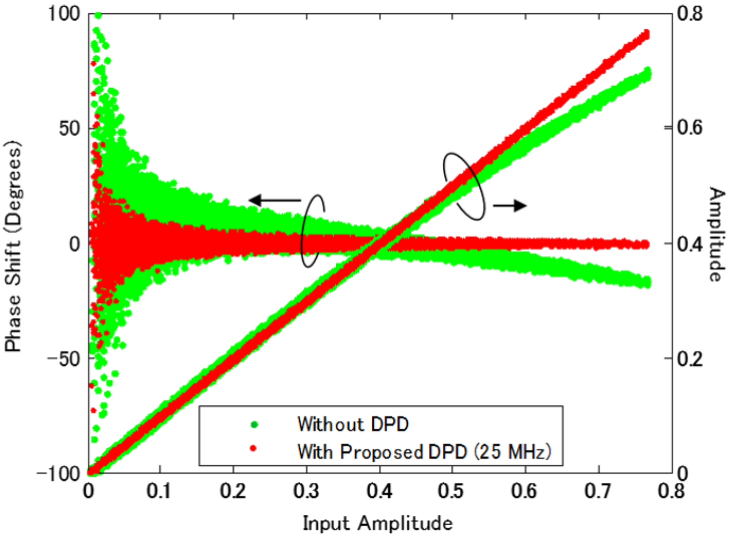
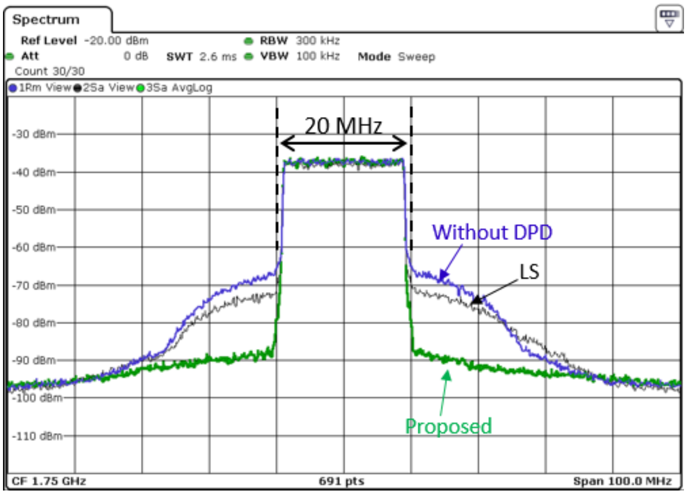


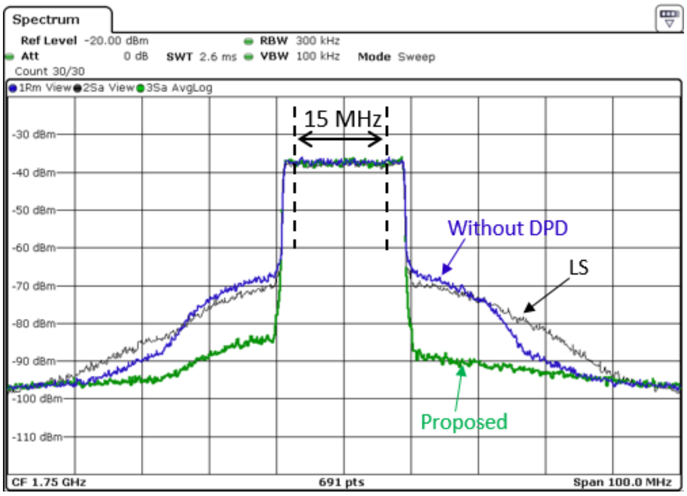
**Fig. 6.5 AM-AM and AM-PM plots with and without DPD (with 25 MHz band-limiting filter) applying LTE signal.**

represents a 10 dB backoff from the saturation point. The PA output power was maintained after the DPD is applied, in order to fairly assess the performance of the DPD.

The performance of the proposed method is first compared with conventional DPDs and the results are reported in Fig.6.4. The test was taken place in three cases: 1) LS algorithm without band-limiting filter; 2) LS algorithm with band-limiting filter; 3) Proposed method with band-limiting filter. In the first case, the acquisition bandwidth is 160 MHz, which is enough to capture the 7th order intermodulation distortion as well as the memory effect. For the followed two cases, the acquisition bandwidth is set to 25 MHz. It can be inspected from the Figure that sufficient performance can be achieved with LS algorithm when the feedback signal is not band-limited. Its performance significantly degraded if band-limiting filter was



(a) With 20 MHz band-limiting filter



(b) With 15 MHz band-limiting filter

**Fig. 6.6 PA output spectra under different band-limiting filters.**

added. The proposed method, on the other hand, can achieve almost the same performance as the LS algorithm with sufficient acquisition bandwidth. The AM-AM and AM-PM plots both with and without DPD are shown in Fig.6.5, where the acquisition bandwidth was set to 25 MHz. It is easy to see that the proposed DPD can suppress both the static nonlinearity and memory effect.

The spectral plots of the PA output with DPD under different band-limiting filters are demonstrated in Fig.6.6. The tests were conducted by assuming the acquisition bandwidth of 20 MHz and 15 MHz respectively. The spectral plots of the PA output with conventional LS and proposed method are also shown in the figure for comparison. As the acquisition bandwidth reduced to 15 MHz, which is even less than the original signal bandwidth, conventional LS algorithm failed to linearize the PA. The proposed method can also provide

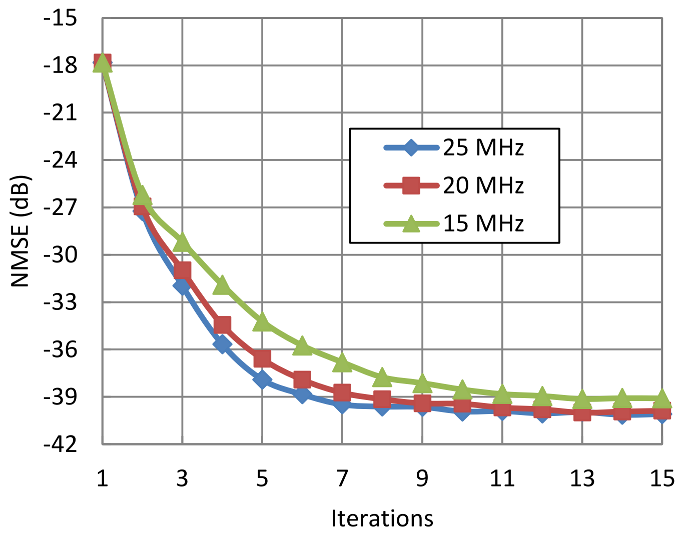


**Fig. 6.7 NMSE at each iterative step for various acquisition bandwidths.**

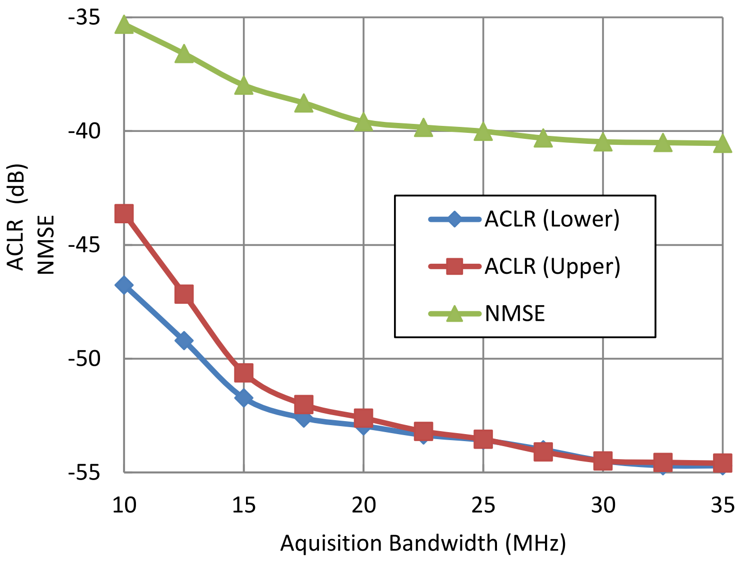


**Fig. 6.8 The ACLR and NMSE performance of the proposed method under various acquisition bandwidths.**

satisfactory performance, even if very narrow acquisition bandwidth is assumed.

The normalized mean square error (NMSE), which was obtained by comparing the input signal after CFR and PA output signal, for each iterative step is illustrated in Fig.6.7. As expected, it shows faster convergence rate and lower NMSEs, when the acquisition bandwidth is 25 MHz. The ACLR and NMSE performance of the proposed method under various acquisition bandwidths are reported in Fig.6.8. As can be inspected from the figure, both

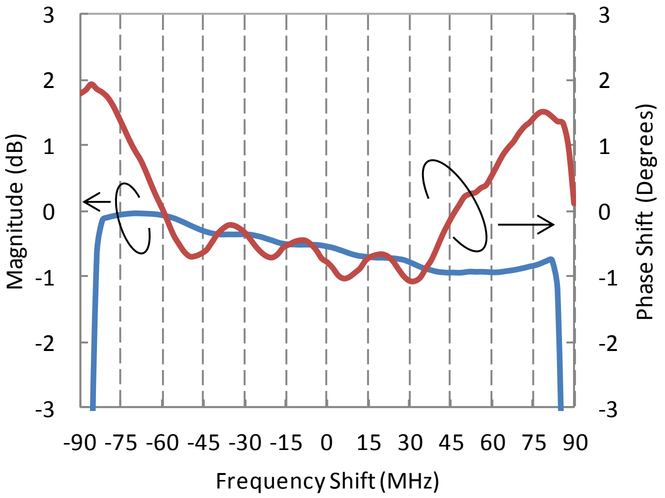


Fig. 6.9 The frequency response of acquisition path.

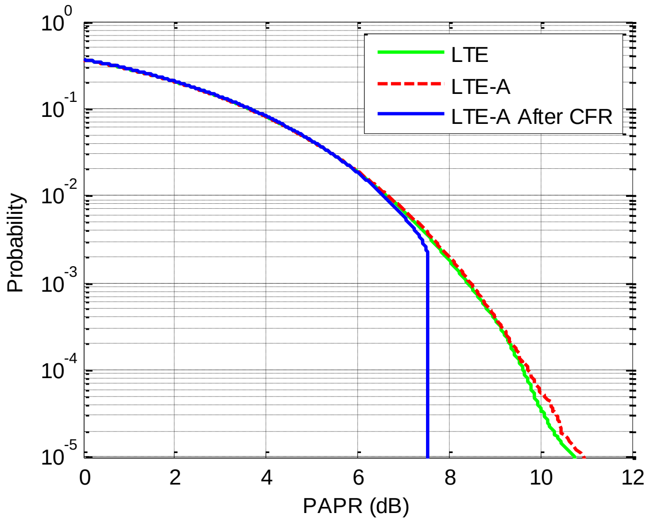
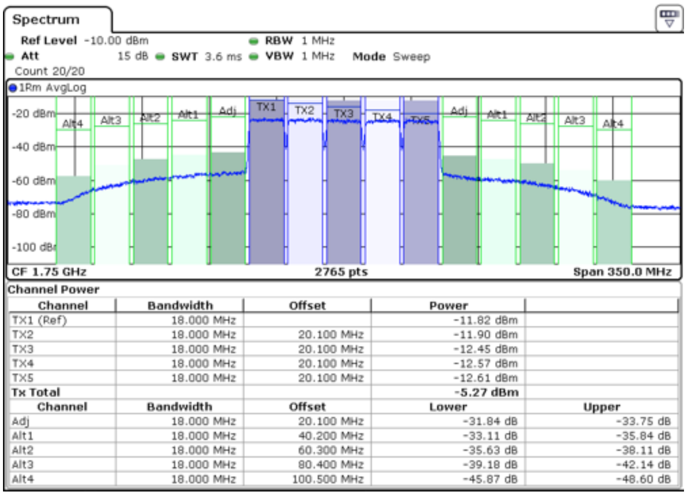


Fig. 6.10 CCDFs for LTE signal, LTE-Advanced signals before and after CFR.

ACLR and NMSE are small if the acquisition bandwidth is larger than 20 MHz. The performance degrades as narrower filter is applied, but it is still acceptable for practical applications. Imbalanced ACLR for upper and lower bands can be inspected when the acquisition bandwidth is smaller than the original signal bandwidth. This is not surprising: the extrapolation works well near the observation band, while the extrapolated data way from the observation band is less reliable. Accordingly, with smaller acquisition bandwidth, spectral extrapolation produces more errors for the extrapolated data.

### 6.4.3 Performance with LTE-Advanced Signal

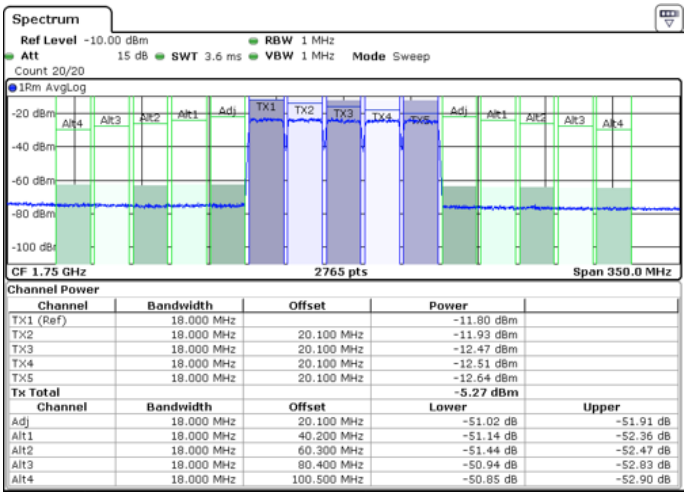


Spectrum
Ref Level -10.00 dBm
Att 15 dB
SWT 3.6 ms
RBW 1 MHz
VBW 1 MHz
Mode Sweep
Count 20/20
1Rm AvgLog
CF 1.75 GHz
2765 pts
Span 350.0 MHz

Channel Power

| Channel | Bandwidth | Offset | Power | |
|---|---|---|---|---|
| TX1 (Ref) | 18.000 MHz | | -11.82 dBm | |
| TX2 | 18.000 MHz | 20.100 MHz | -11.90 dBm | |
| TX3 | 18.000 MHz | 20.100 MHz | -12.45 dBm | |
| TX4 | 18.000 MHz | 20.100 MHz | -12.57 dBm | |
| TX5 | 18.000 MHz | 20.100 MHz | -12.61 dBm | |
| **Tx Total** | | | **-5.27 dBm** | |
| **Channel** | **Bandwidth** | **Offset** | **Lower** | **Upper** |
| Adj | 18.000 MHz | 20.100 MHz | -31.84 dB | -33.75 dB |
| Alt1 | 18.000 MHz | 40.200 MHz | -33.11 dB | -35.84 dB |
| Alt2 | 18.000 MHz | 60.300 MHz | -35.63 dB | -38.11 dB |
| Alt3 | 18.000 MHz | 80.400 MHz | -39.18 dB | -42.14 dB |
| Alt4 | 18.000 MHz | 100.500 MHz | -45.87 dB | -48.60 dB |

(a) Without DPD



Spectrum
Ref Level -10.00 dBm
Att 15 dB
SWT 3.6 ms
RBW 1 MHz
VBW 1 MHz
Mode Sweep
Count 20/20
1Rm AvgLog
CF 1.75 GHz
2765 pts
Span 350.0 MHz

Channel Power

| Channel | Bandwidth | Offset | Power | |
|---|---|---|---|---|
| TX1 (Ref) | 18.000 MHz | | -11.80 dBm | |
| TX2 | 18.000 MHz | 20.100 MHz | -11.93 dBm | |
| TX3 | 18.000 MHz | 20.100 MHz | -12.47 dBm | |
| TX4 | 18.000 MHz | 20.100 MHz | -12.51 dBm | |
| TX5 | 18.000 MHz | 20.100 MHz | -12.64 dBm | |
| **Tx Total** | | | **-5.27 dBm** | |
| **Channel** | **Bandwidth** | **Offset** | **Lower** | **Upper** |
| Adj | 18.000 MHz | 20.100 MHz | -51.02 dB | -51.91 dB |
| Alt1 | 18.000 MHz | 40.200 MHz | -51.14 dB | -52.36 dB |
| Alt2 | 18.000 MHz | 60.300 MHz | -51.44 dB | -52.47 dB |
| Alt3 | 18.000 MHz | 80.400 MHz | -50.94 dB | -52.83 dB |
| Alt4 | 18.000 MHz | 100.500 MHz | -50.85 dB | -52.90 dB |

(b) With proposed DPD

**Fig. 6.11 Test PA spectrum which is an LTE-Advanced signal with 5 contiguous carrier components. The acquisition bandwidth is 160 MHz.**

The results obtained above can provide us sufficient confidence for developing a wideband DPD for LTE-Advanced with a narrow acquisition path and low speed ADC. In wideband applications, however, the frequency response of the band-limiting filter should be taken into account, as the acquisition path cannot be ensured to be flat in the bandwidth of interest. The equivalent baseband frequency response of the acquisition path, which was obtained by a multi-tone test, is shown in Fig.6.9. We can see the magnitude fluctuation of the feedback filter is around 1dB, and meanwhile the phase ripple is more than 3 degrees for the center 160 MHz band. The compensation filter was obtained corresponding to the measured filter response.

The test signal was an LTE-Advanced signal with 5-CCs of 20 MHz bandwidth. It is worth to mention that the channel spacing of the CCs should be multiple times of 300 KHz

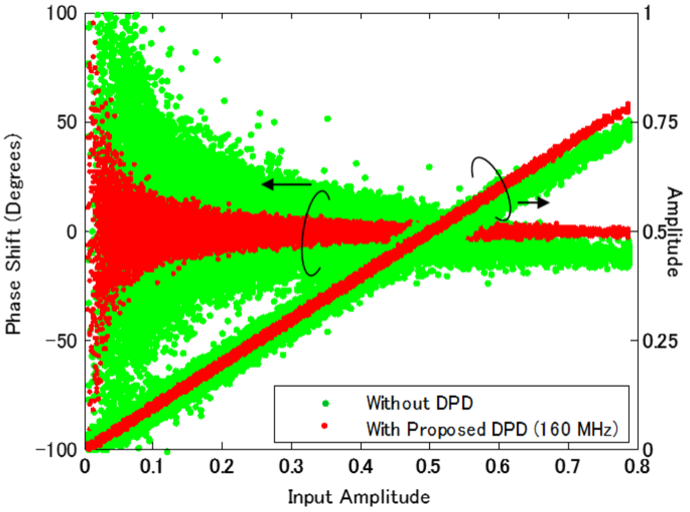


**Fig. 6.12 AM-AM and AM-PM plots with and without DPD, where the signal is an LTE-Advanced signal with five CCs.**

to ensure the orthogonality of the subcarriers with 15 KHz spacing, and hence 20.1 MHz was set in our test. This results in a contiguous signal bandwidth more than 100 MHz. Note that carrier aggregation does not further increase the PAPR of the signal, and this can be clearly seen from the complementary cumulative distribution function (CCDF) plots shown in Fig.6.10. The PAPR for LTE-Advanced signal can be reduced and strictly kept under 7.5 dB with CFR as shown in this figure. Fig.6.11(a) shows the PA output spectral plot without DPD. The maximum driving condition has been set so as to the highest peak reaches the 3 dB compression point. The resulting ACLRs at 20.1 MHz offset were around -32.41/-30.95 dB for lower and upper bands, as can be seen from the figure. The subsequent plot shown in Fig.6.11(b) demonstrates the effectiveness of the proposed technique facilitating linearization of the PA. The ACLRs at 20.1 MHz offset were improved by over 20 dB for both upper and lower bands. It is noticed that the spectrum of the proposed method is already close to the noise floor of the measurement setup. The AM-AM and AM-PM plots are shown in Fig.6.12. The characteristics without DPD are very dispersive, because the band-limiting filer has eliminated part of the frequency components of the feedback signal and this introduces substantial linear distortions. Fig.6.12 also reveals linearized AM-AM and AM-PM characteristics with the proposed DPD, from which it can be seen that the dispersion is significantly reduced.

The involved EVMs for respective CCs, which were obtained by measuring the signal of one time slot, are tabulated in Table I. Since EVM is a measure to quantify the performance of the whole system, the EVM of the test bed (directly connecting the two instruments) is also shown in the table. Note that the errors are mainly introduced by the feedback filter. The EVM can much improved if the compensation filter is applied. Furthermore, the EVM was also degraded by the CFR, as result of in-band distortion. The validity of the proposed DPD technique can be seen by comparing the EVM performance with and without DPD.

The results of the tests employing LTE-Advanced signal of two inter-band aggregated

**Table 6-1 The Measured EVM for each CC**

| | *CC*1 | *CC*2 | *CC*3 | *CC*4 | *CC*5 |
|---|---|---|---|---|---|
| Test Bed | 1.67% | 1.45% | 1.42% | 1.42% | 1.63% |
| Test Bed with Compensation Filter | 0.32% | 0.29% | 0.25% | 0.27% | 0.34% |
| CFR Only | 2.02% | 2.01% | 2.01% | 2.02% | 2.00% |
| PA Only | 9.89% | 9.44% | 9.01% | 9.16% | 9.85% |
| PA With CFR & DPD | 2.99% | 2.71% | 2.69% | 2.64% | 2.81% |

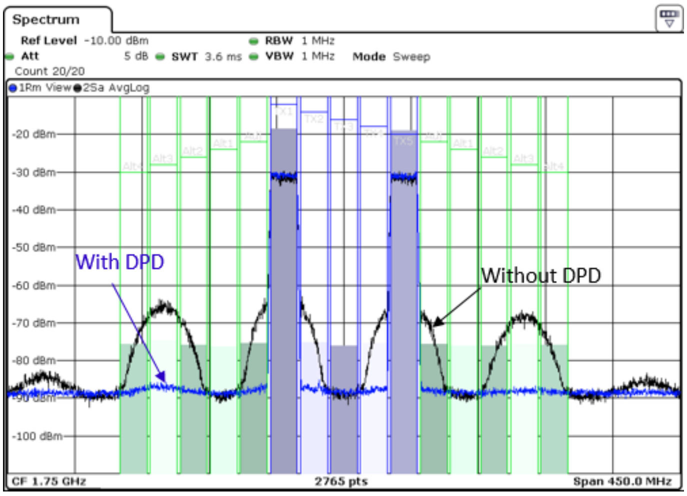


**Fig. 6.13 The spectral plots of the PA output with and without DPD, where the signal is a two-carrier LTE signal in which the carrier space is 80.4 MHz.**

CCs are illustrated in Fig.6.13. The two CCs were separated by three channels (i.e. 80.4 MHz). The nonlinear distortion, as can be seen from the figure, has larger impact on the 3rd alternative channel than the adjacent channel. With proposed DPD, the intermodulation distortion can be efficiently suppressed.

A signal with seven LTE carrier components, which represents a contiguous spectrum of 140 MHz bandwidth, was also applied in our test. Note that although such signal may not exist in real communication systems, it is still desirable to assess the potential of the proposed DPD for PA linearization applying very wideband signal. The results demonstrated in Fig.6.14 show that the ACLR at 20.1 MHz offset reached -51.71 dBc and -51.80 dBc for lower and upper band, respectively, with the proposed DPD. The results obtained by applying signal of 160 MHz bandwidth are reported in Fig.6.15. It can be found that the out-of-band emission was significantly suppressed, even if the bandwidth of the original signal

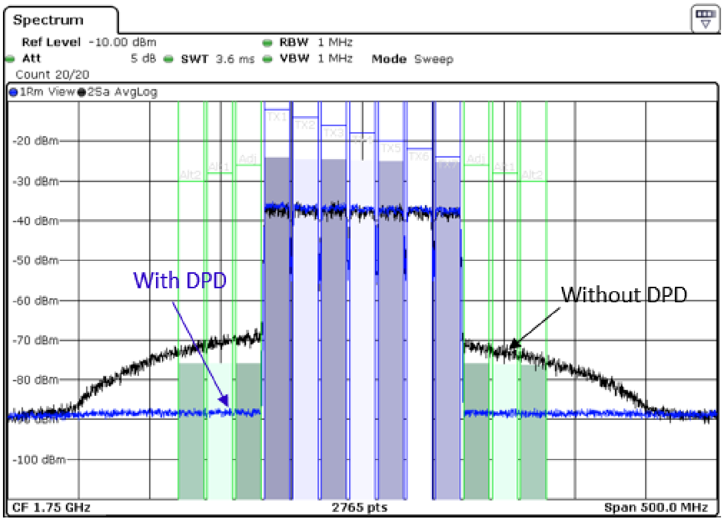


**Fig. 6.14 The spectral plots of the PA output with and without DPD, where the signal is comprised of seven CCs, which represents a contiguous spectrum with140 MHz bandwidth.**

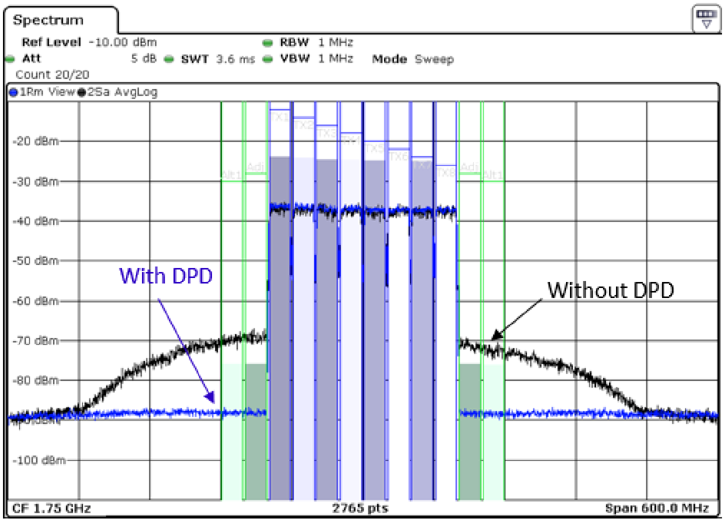


**Fig. 6.15 The spectral plots of the PA output with and without DPD, where the signal is comprised of eight CCs, which represents a contiguous spectrum with 160 MHz bandwidth.**

has reached the bandwidth limit of the acquisition path.

Using a wider band receiving path, i.e. a FSW with 300MHz IF bandwidth, linearization of very wideband signal can be accomplished, as shown in Fig. 6.16, where the signal is comprised of 16 LTE 20MHz CCs to form a 320MHz contiguous spectrum. It is not difficult to see that the proposed technique works well for dealing with such a wideband signal, and this is the widest DPD results have ever been reported.

The next test used intra-band carrier aggregated LTE-Advanced signal with 1 CC in the lower side band and 3 CCs in the upper side band. The two bands are spaced by 200 MHz, or namely with 9 unoccupied 20 MHz channels, as shown in Fig. 6.17. One can see that the nonlinear PA causes both near-band distortion as well as intermodulation distortion away

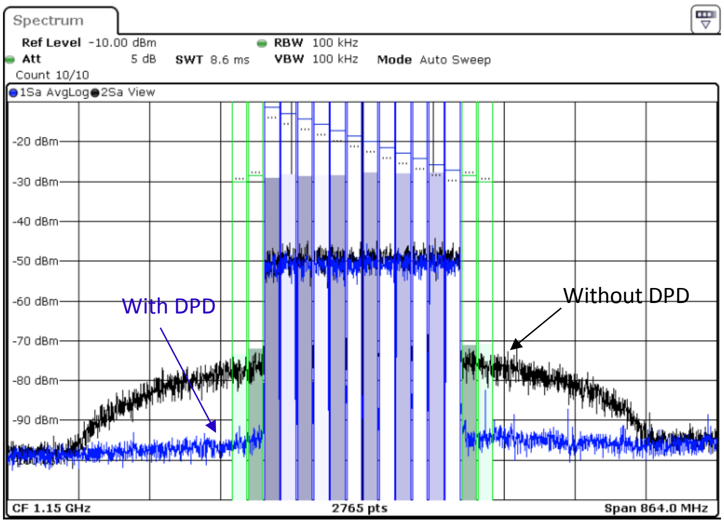


**Fig. 6.16** **The spectral plots of the PA output with and without DPD, where the signal is comprised of 16 CCs, which represents a contiguous spectrum with 320 MHz bandwidth.**

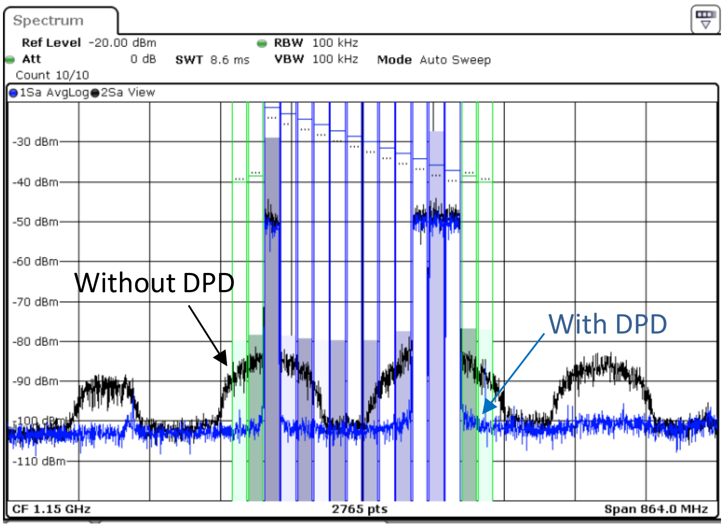


**Fig. 6.17** **Spectral plots of the PA output with and without DPD, where the signal is an intra-band aggregated LTE-Advanced signal with 1 CC in lower sideband 3 CCs in the upper side band.**

from the bands. With SE-DPD, the obvious improvement on the both types of distortion is obtained. Outband radiation spectrum level is -50 dBc for this case.

The last test is conducted with 5 CCs located in the upper side band. This results in a spectrum with unequal out-of-band distortion for upper and lower bands, as shown in Fig.6.18. The distortion can be significantly reduced by the use of SE-DPD.

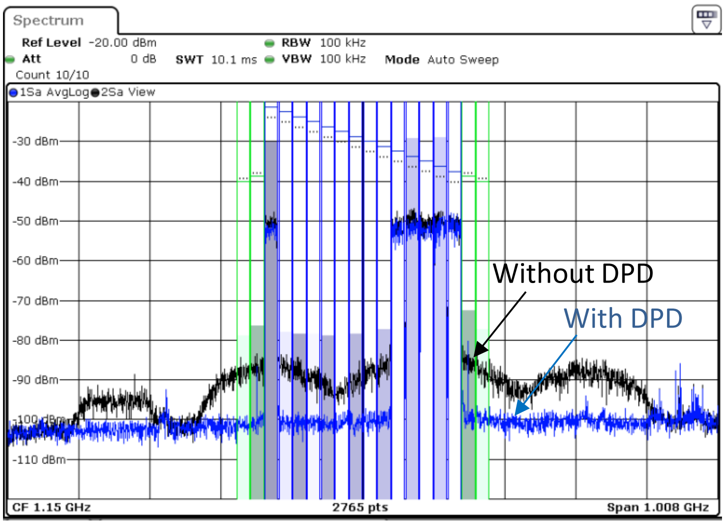


**Fig. 6.18** Spectral plots of the PA output with and without DPD, where the signal is an intra-band aggregated LTE-Advanced signal with 1 CC in lower sideband 5 CCs in the upper side band.

# Chapter 7.

# Blind Nonlinear Compensation for Receiver RF Front-End

A novel nonlinear compensation technique for RF (radio frequency) receiver front-end linearization is presented in this chapter. Compared with the frequency-domain blind method, the proposed technique has lower complexity and allows the compensation to work in fully real time. The adaption criteria employed in the compensator is aiming to minimization of the out-of-band intermodulation distortion components. To enhance the performance of the compensator, de-embedding is employed to reconstruct the polynomial model so that the contribution of the linear component is minimized in the nonlinear functional. The proposed technique has low complexity and allows the compensator to operate in real time. The experimental results indicate that the proposed compensation technique can effectively suppress distortion due to receiver front-end nonlinearity.

## 7.1 PROBLEM OF ADJACENT CHANNEL INTERFERENCE

As the emergency of new wireless communication standards and diversified services for end users, radio environment is becoming more complex in the context of frequency spectrum and density of radio equipment. Thus multi-mode mobile terminals, which must support multiple radio systems, are facing radio coexistence issues. Most of the coexistence issues are attributed to the nonlinearity of the receiver front-end. When the receiving signal is weak, it is negatively affected by the strong signal from another transmitter.

The most noticeable issue appears as reduced sensitivity due to noise level increase in the receiving band. As depicted in Fig.7.1, it is caused by the interferer’s out-of-band radiation that is the result of intermodulation (IM) at the receiver front-end. Although an RF (radio frequency) band-pass filter can be used to reduce the power of interference, it has limited attenuation and cannot suppress the noise in the receiving band.

This chapter proposes a novel compensator that employs baseband signal processing. The input signal and interfering signal are assumed unknown and thus the blind identification algorithm is applied to achieve real-time adaptive compensation. The bandwidth and center frequency of the interference are only required as *a priori* knowledge so that the intermodulation components can be clearly defined. The proposed compensator has been implemented in Xilinx Kinttex-7 field programmable gate array (FPGA), and experimental

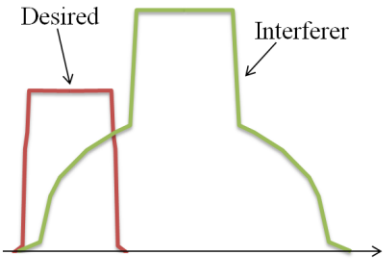


**Fig. 7.1 Intermodulation distortion of strong interference corrupts the de-sired signal.**

results are presented to show its effectiveness.

## 7.2 FREQUENCY-DOMAIN BLIND NONLINEAR COMPENSATION ALGORITHM

The blind nonlinear compensation algorithm for band limited signal was first proposed in [128] and it was also applied by the author in [129] for the compensation of a software-defined receiver. Here we only consider the compensator a memoryless polynomial of $P$th-order, and the distorted signal $y$(n) is fed to this polynomial so that the output is given by

$$s(n) = [y(n), \mathbf{u}(n)]\mathbf{A} \tag{7.1}$$

where

$$\mathbf{y}(n) = [y^2(n), y^3(n), \cdots, y^P(n)] \tag{7.2}$$

is the data matrix of the polynomial, and $\mathbf{A}=[\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_P]^{\mathrm{T}}$ is the coefficient vector for the polynomial. Note that this constraint assumes $\alpha_1=1$, which indicates that the desired component has a gain of 1. The scaling that $\alpha_1 \neq 1$ is also applicable since the constant scaling can be easily restored. The reason why $\alpha_1$ is fixed is to prevent the signal from extinction: one solution for minimizing the out-of-band component is $\alpha_i=0$ ($i=1,2,\ldots,P$). Then an N-point DFT (discrete Fourier transform) is applied to $s(n)$, which is given by

$$\mathbf{Y}(k) = \mathbf{W}_N \mathbf{y}(n) \tag{7.3}$$

where $\mathbf{W}_N$ is an $N$ by $N$ DFT transform matrix. Since the objective of the nonlinear compensation is to eliminate the out-of-band emission, the cost function can be defined as the out-of-band signal

$$J(\mathbf{A}) = \left\| [\hat{Y}(k) + \mathbf{Y}(k)]\mathbf{A} \right\|_2^2 \tag{7.4}$$

where the hat denotes the corresponding out-of-band part. Minimization of $J(\mathbf{w})$ against $\mathbf{A}$ yields the estimation of the compensator parameters that negate the out-of-band energy. With (7.4), both least squares solution and iterative algorithm can be developed.

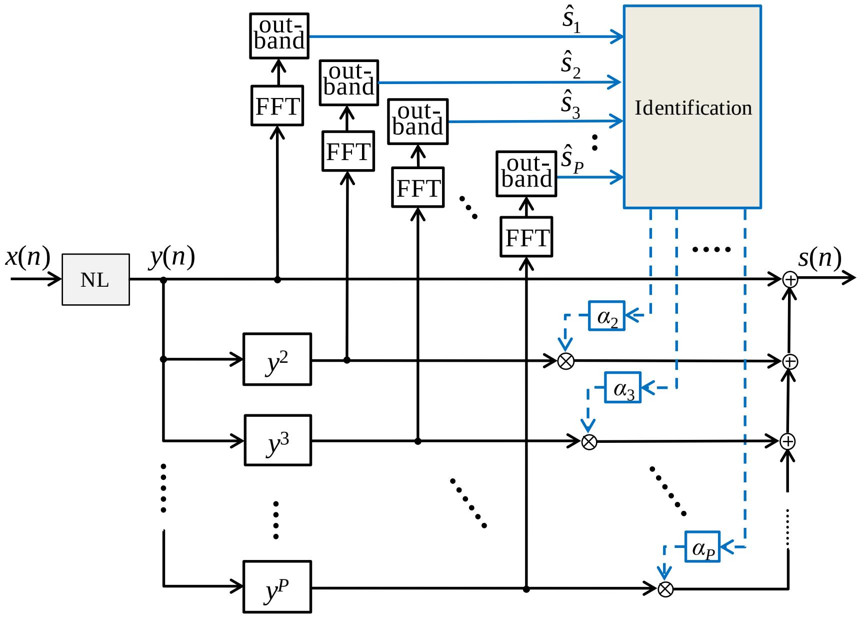


**Fig. 7.2 The diagram of the frequency domain blind nonlinear compensation.**

It can be seen from the foregoing presentation that it requires $P$-FFT (fast Fourier transform) blocks to transform the monomials to frequency domain. More blocks may be required as more complicated model is employed. The simplified diagram describing the algorithm is graphically shown in Fig.7.2.

Another aspect of the algorithm is that the FFT is a block-by-block operation, which requires buffering the signal stream. As consequence the update of the parameters cannot be completely in sample-by-sample manner, hindering fully real time operations.

## 7.3 NONLINEAR COMPENSATOR WITH TIME DOMAIN BLIND IDENTIFICATION ALGORITHM

The compensation scheme proposed in this chapter is illustrated in Fig.7.3. The principle is to employ a filter to obtain the IMD components excluding the bands of the desired signal and interferer. The filter output is utilized as the minimization cost function, and the coefficients of the compensator are adaptively updated. The block $\boldsymbol{\tau}$ denotes a sampling delay buffer which is to align the input signal with the filtered output. Since the filter is predetermined before implementation, $\boldsymbol{\tau}$ is a fixed value. The compensator is polynomial based and its detail will be discussed in the following sections.

### 7.3.1 The Compensator Model

The most common tools for nonlinear modeling and compensation are polynomial-based models. If only memoryless nonlinearity is assumed, the baseband polynomial as given be

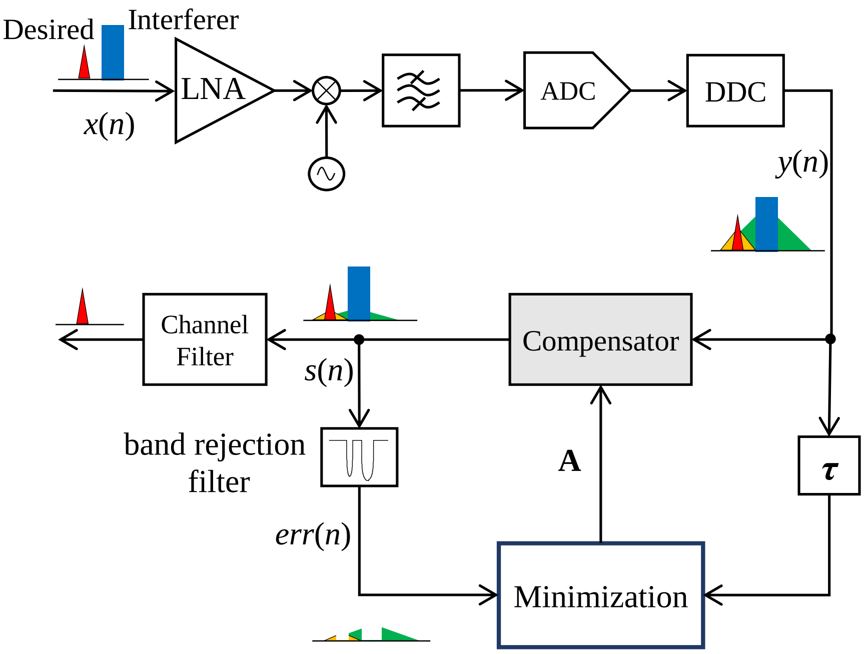


**Fig. 7.3 Block diagram of the blind nonlinear compensation scheme.**

low can be used as the compensator model

$$s(n) = y(n) + \sum_{i=2}^{N} a_i \left| y(n) \right|^{i-1} y(n) \tag{7.5}$$

where $N$ is the nonlinear order. Note that this model assumes $a_1=1$, which indicates that the linear term has a gain of 1. Note that scaling of $w_1 \neq 1$ is also possible because the constant scaling can be easily restored. The reason for fixing $w_1$ is to prevent the signal from extinction. Actually, a global solution for out-of-band component minimization is $a_i=0$ ($i$=1,2,…,N), so that $a_1$ is fixed to prevent the adaption from converging to this solution. However, a side effect arises that fixing $w_1$ will have a negative impact on the estimation of the coefficients, in terms of such as slow convergence rate, since the rest of the functions (monomials) have strong correlation with $y(n)$.

An important manifestation of nonlinear distortion is the occurrence of out-of-band emission. Hence minimization of this signal is a nature consequence of nonlinear compensation. Nevertheless, it is frequently missed by researchers that the nonlinear functional in the compensator generates not only the out-of-band emission, but also linear component of much higher power than the intermodulation component. For instance, if we take a closer look at the third order monomial $x^3$, where $x$ is assumed to be a two tone signal given by $x=\cos(\omega_1 t)+\cos(\omega_2 t)$, there yields

$$\begin{aligned} x^3 &= \left[\cos(\omega_1 t) + \cos(\omega_2 t)\right]^3 \\ &= \cos^3(\omega_1 t) + \cos^3(\omega_3 t) + 3*\cos^2(\omega_1 t)\cos(\omega_2 t) + 3*\cos^2(\omega_2 t)\cos(\omega_1 t) \end{aligned} \tag{7.6}$$

If only the components near the frequency of interest are considered, it gives

$$x^3\Big|_{\omega\approx\omega_1,\omega_2} = 2\cos(\omega_1 t)+2\cos(\omega_2 t)+ \frac{3\cos(2\omega_2 t-\omega_1 t)}{4}+\frac{3\cos(2\omega_1 t-\omega_2 t)}{4} \tag{7.7}$$

The above calculation shows that the power for the IM3 (3rd-order intermodulation) components is lower than the in-band signal by around 8.5dB. Similar results can also be obtained for other nonlinear functional. Therefore, it can be concluded that the nonlinear functions (e.g. the monomials) implicitly incorporates the linear components which dominates the output power.

Based on the foregoing discussion, it is reasonable to de-embed the linear component from the nonlinear functional, so that the pure intermodulation components can be utilized. De-embedding can be accomplished by simply subtracting the input signal from the nonlinear functional, which yields the model as illustrated in Fig.7.4, and the expression is given by

$$s(n) = y(n)+\sum_{i=2}^{N} a_i\phi_i(n) \tag{7.8}$$

where

$$\phi_i(n) = |y(n)|^{i-1}\, y(n)+\beta_i y(n) \tag{7.9}$$

for which

$$E[\phi_i(n)y^*(n)] = 0 \tag{7.10}$$

should be met. *E*[ ] is the expectation and the superscript * denotes conjugation. Therefore the parameters $\beta_i$ (i=2, 3, … , *N*) be determined with the following equation

$$\beta_i = \frac{E[|y(n)|^{i+1}]}{E[|y(n)|^2]} \tag{7.11}$$

With this structure, the nonlinear functional is explicitly uncorrelated with the linear term, so that the contribution of linear component to the nonlinear functional is minimized. This has many benefits such as faster convergence rate and lower level requirement on the out-of-band filter.

Subjecting to fixed-point processing, direct implementation of (7.8) has high complexity and potential numerical stability problems because many multiplications are involved. It is easy to see from (7.4) that the baseband polynomial can be replaced by the multiplication of the signal and a lookup table (LUT) indexed by the signal amplitude, if respective exponentials are stored in the LUT. As the LUT can represent any desired equations, the function of baseband polynomial even up to infinite order can be firmly imitated. The magnitude computation can be accomplished with a CORDIC algorithm [115].

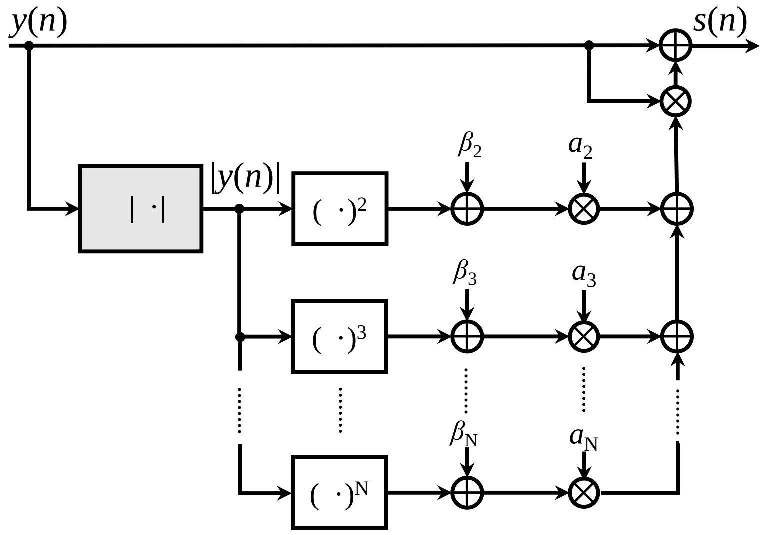


**Fig. 7.4 Proposed compensator model.**

### 7.3.2 Blind Nonlinear Compensation Algorithm

Due to simplicity and feasibility of hardware implementation, least mean square (LMS) is applied for the adaptively updating the compensator. We first define the data matrix for the compensator $\Psi(n)=[\varphi_2(n), \ldots, \varphi_N(n)]$, so that the compensator output can be expressed by

$$s(n) = y(n) + \Psi(n)\mathbf{A} \tag{7.12}$$

where $\mathbf{A}=[a_2, a_3, \ldots, a_N]^T$ is a column vector containing the coefficients.

The 2-norm of the filter output, which is the out-of-band energy, is defined as the cost function:

$$J(\mathbf{A}) = \|err(n)\|_2^2 \tag{7.13}$$

Herein

$$err(n) = h(n) * [\Psi(n)\mathbf{A}] \tag{7.14}$$

where $h$(n) is the impulse response of the out-of-band filter. The resulting gradient estimate is given by

$$\frac{\partial J(\mathbf{A})}{\partial \mathbf{A}} = -2h^*(n) * \Psi^H(n) err(n) \tag{7.15}$$

where the superscript $^H$ denotes the Hermite transposition. It can be seen from the expression in (7.15) that directly applying the LMS algorithm requires filtering the data matrix. To reduce the complexity of the algorithm, the following approximation is introduced.

$$\frac{\partial J'(\mathbf{A})}{\partial \mathbf{A}} = -2\Psi^H(n) err(n) \tag{7.16}$$

Finally, the resulting LMS can be developed, whose updating equation is

$$\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} - \gamma\ \Psi^{H}(n-\tau)err(n) \tag{7.17}$$

where the data matrix is delayed by $\gamma$, which is time aligned with the filter output, see Fig.7.3. Other gradient based algorithms, such as RLS (recursive least square), are also applicable for the proposed compensator. However, they are not discussed here because of the length limit of this thesis.

### 7.3.3 Band Rejection Filter

The filter used in the proposed compensation technique is to eliminate the desired signal and interference and to obtain out-of-band signal. Since the compensation processing is in baseband, the desired signal is located in the center in frequency domain while the interference is at the side of it. Therefore, the frequency response of the filter should be asymmetric, resulting in a complex coefficient filter. Thus, implementation of the out-of-band filter requires three real-variable filters.

Another aspect of the filter design here is that the filter has no standard frequency response like low-pass, high-pass and band-pass. The specific application in the proposed compensator needs design of a filter with arbitrary response. The design methodologies have been studied extensively by many researchers, and can be found in many literatures. Here, chebyshev approximation method [130] was employed to calculate the coefficients of the desired filter.

## 7.4 PERFORMANCE EVALUATION

The proposed compensator has been implemented in FPGA with an external clock of 200MHz. A series connection of RF amplifiers providing a gain of over 36dB and peak power around 25dBm at 2.4 GHz band was used as the LNA. A vector signal analyzer (VSA) with 110MHz acquisition bandwidth, which is an RSA6100a from Tektronix, was employed to digitize and convert the signal to baseband. Then the signal was uploaded to the FPGA and the compensated signals were obtained with the Chipscope tool.

The following signals and parameters were employed in the measurements.

- The desired signal was an LTE (long term evolution) signal featuring 16-QAM for all user subcarriers, 2347.5 MHz center frequency, 4.5 MHz occupied bandwidth, and PAPR (peak-to-average power ratio) of 9.84 dB @99.99% probability.
- The interference signal was also an LTE signal featuring 2355 MHz center frequency, 9 MHz occupied bandwidth, and PAPR (peak-to-average power ratio) of 10.13 dB @99.99% probability.

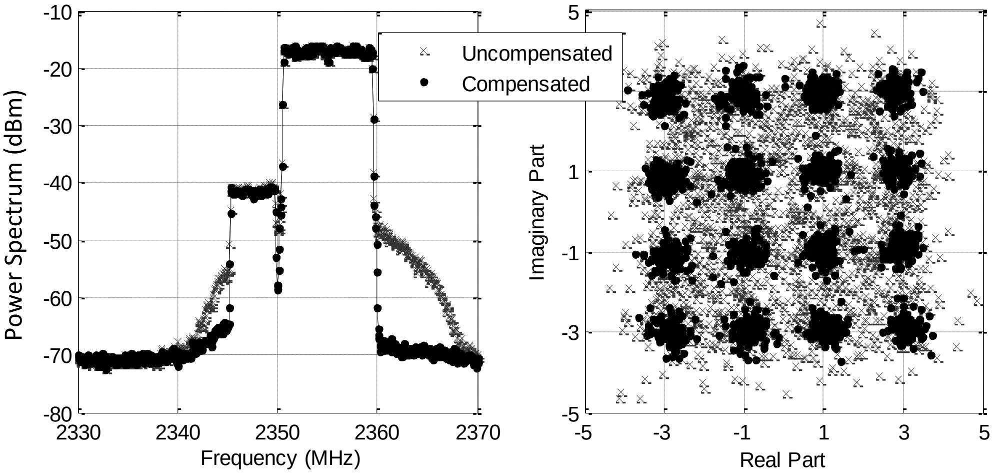


**Fig. 7.5 Spectral plots and constellations for the LNA output with and without the compensator, where the power of the desired signal is -50dBm and the blocker power is -21 dBm.**

- The compensator model applies a polynomial of 5th order, the step size $\gamma$ for iteration is set to 0.5.

To analyze basic performance of the proposed compensator, a set of simulation was carried out in the presence of AWGN (additive white Gaussian noise). The spectral plots as well as the constellations shown in Fig.7.5 were obtained assuming the respective desired signal power and interferer power as -50dBm and -21dBm. Note that the interfering signal backoff here is close to the PAPR of the LTE signal which is about 10 dB, and the peak power is just below the saturation point. The intermodulation distortion components generated from the strong interferer is in a level comparable with the desired signal, as can be inspected from the spectrum.

The resulting constellation is scattered by the adjacent channel interference. With the proposed compensator, the intermodulation distortion can be greatly suppressed so that acceptable constellation of user data can be achieved.

To assess the in-band error performance, error vector magnitude (EVM) for all the user subcarrier was measured. The EVM results obtained by sweeping the interferer power are demonstrated in Fig.7.6, where the desired signal power is set to -50, -45 and -40 dBm, respectively. With larger interferer power and likewise the intermodulation power falling on the desired signal band, the EVM becomes worse. The EVM can be greatly improved with the proposed compensator.

The EVM performance also depends on the desired signal power, and this is shown in Fig.7.7 where the desired signal power was swept. The interfering signal power $P_b$ were set to -21, -22 and -23 dBm and three of these cases assumed the backoff close to the PAPR so that substantial nonlinear distortion occur.

The EVM performances degrade in the low power region in Fig.7.7. It is mainly due to the reduced SNR in the VSA. The quantization noise introduced by the analog to digital

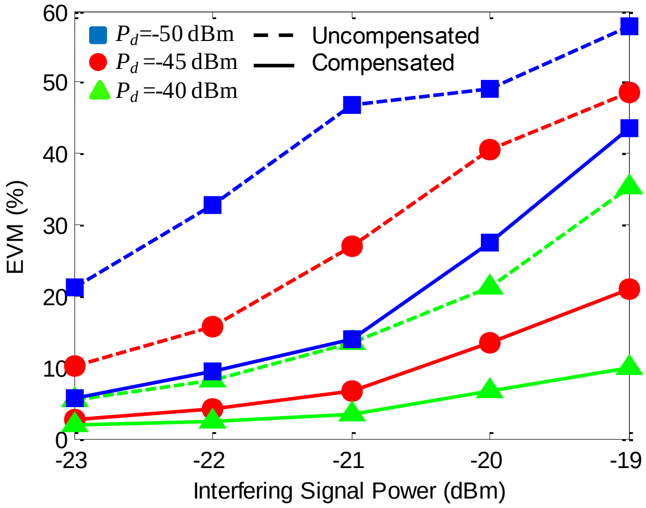


**Fig. 7.6 Measured EVMs with interfering signal power swept therein the desired signal power is set to -50, -45 and -40 dBm, respectively.**

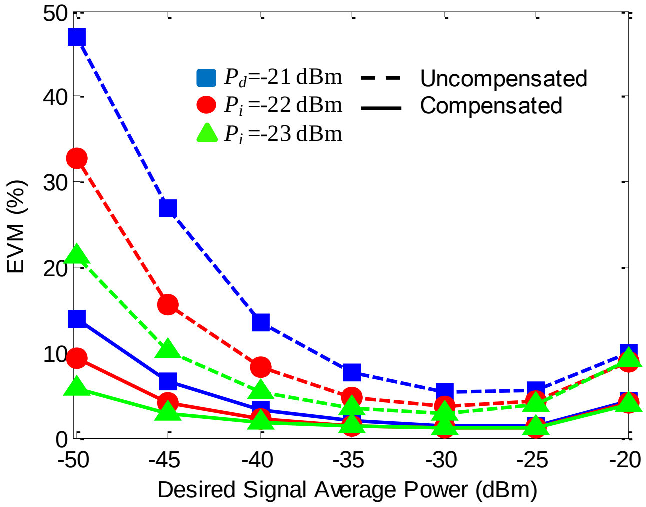


**Fig. 7.7 Measured EVMs with desired signal power swept.**

conversion in the VSA has a negative impact to the measurement, especially when the desired signal power is low.

On the other hand, EVM degrades along with the increase of desired power when it is more than -25 dBm. IM distortion occurs in this region and it dominates the whole distortion power.

# Chapter 8.

# Adaptive Compensation of Inter-Band Modulation Distortion for Tunable Concurrent Dual-Band Receivers

For concurrent dual-band receiver with shared radio frequency (RF) front-ends, the received signals are prone to be contaminated by the inter-band modulation components even if the signals are located in distant frequency bands. A novel signal processing technique for blind compensation of nonlinear distortion in concurrent dual-band receiver is presented in this chapter. To enhance the performance of the compensator, de-embedding is employed to reconstruct the polynomial model. The compensator coefficients updating is accomplished in a block-by-block manner, such that the computing complexity and adaption speed can be well balanced. Finally, the performance of the compensator is validated by the results obtained from actual measurements.

## 8.1 THE NONLINEAR DISTORTION OF CONCURRENT DUAL-BAND RECEIVER

The front-ends for next generation wireless communication system are becoming very complex with over 10 different bands ranging from 700 MHz to 2700MHz. Therefore, a number of filters, antennas, amplifiers and transceivers are required to support respective bands that are switched depending on the demands. On the other hand, recent explosion of mobile traffic, mainly stimulated by smart phones, forces simultaneous use of multiple bands and protocols. One example of this is the carrier aggregation technique [49]-[51] employed in the LTE-Advanced system for improving the overall communication throughput. A possible deployment scheme for inter-band carrier aggregation, for example, is to transmit two or more signals in different bands with physically separated transmitters, as shown in Fig.8.1. A macro base station that works in lower frequency band provides a large coverage while another femto base station or repeater working in higher frequency band is set in traffic hotspots to extend throughputs by another carrier component. In such a scheme,

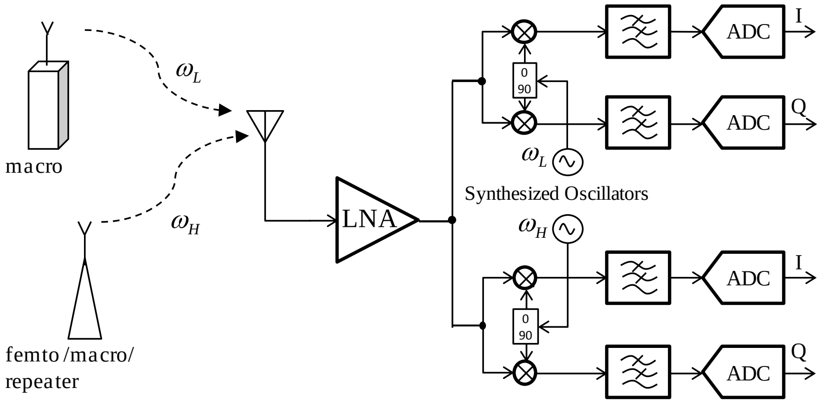


**Fig. 8.1 Tunable concurrent dual-band receiver in inter-band carrier aggregation deployment. The two transmitters are working at the center frequencies of $\omega_L$ and $\omega_H$, respectively.**

the two signals are received and aggregated by a single receiver.

In order to reduce or eliminate the number of RF components and enable efficient use of these circuits, the concurrent multi-band operation receiver [131] has been proposed. With the advances of tunable components [132], receivers that flexibly cover multiple RF bands have been developed. In this chapter, the concurrent dual-band receiver architecture shown in Fig.8.1 is considered. The broadband antenna and low noise amplifier (LNA) allow receiving signals in different bands, and the subsequent synthesized (tunable) local oscillators select and convert the desired signals to base-band via direct down conversion (DDC) structure. Due to its high reconfigurability, this receiver is an attractive candidate technique to realize the software- defined radio (SDR).

On the other hand, the concurrent dual-band operation forces the receiver into a more difficult situation. The received signal strength of respective signals fades independently between two RF bands. Also, separate locations of the two transmitters make the average received power different. When the two signals pass through the front-end of the receiver, the possibility of causing intermodulation due to the strong signal/signals significantly increases. In the case of the concurrent dual-band receiver depicted in Fig.8.1, two kinds of distortion may arise, with their respective impact on the spurious level of the signals, see Fig.8.2. The self-generated distortion is referred to the spectral components stemming from the signal in the same frequency band, while inter-band modulation is the distortion caused by mixing of the two signals in different bands. Inter-band modulation components would be generated and imposed on the two received signals, so long as the two signals are experiencing the same nonlinearity even if they reside in distant bands. The intermodulation distortion falls into the desired signal bands and cannot be simply removed by the subsequent channel filter. Thus, a countermeasure is necessary to avoid severe degradation of the receiver performance caused by nonlinear distortion.

Only a few have considered the inter-band modulation so far [133][134]. However, the work of [133] lacks the implication in practical situations as it assumes idealized synchronization to correctly recover the reference signals. In [134], a more specific case was

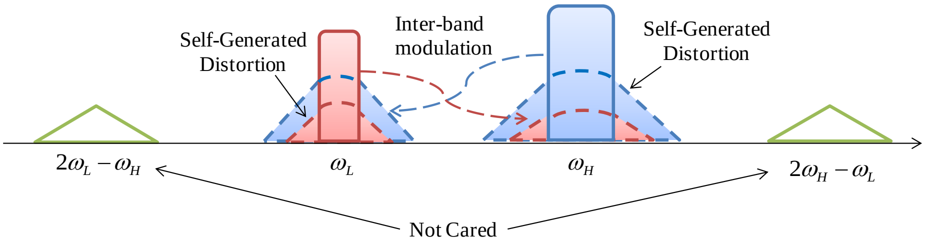


**Fig. 8.2 Nonlinear distortion of concurrent dual-band receiver.**

considered; one of the received signals was transmitted from the same terminal and thus its influence can be accurately removed. As a consequence, both of these works can be seen as special cases of the decision feedback equalizer. Recently, some researchers have studied the predistortion of concurrent dual-band transmitters [135]. At the transmitter side, both the original signal and distorted signal are known, so that the error function can be easily defined by directly comparing the two signals. Nevertheless, the issue addressed in this chapter is different from predistortion because only the nonlinear output signals are available without any information of original signals. Consequently, linearization of the receiver is essentially a blind nonlinear compensation problem.

This chapter focuses on the case illustrated in Fig.8.2 and proposes a corresponding remedial technique. The nonlinear characteristic of the front-end is assumed unknown and thus blind identification algorithm is applied to achieve real-time adaptive compensation. The bandwidths and the center frequencies of the two signals are only required as *a priori* knowledge, so that the intermodulation components can be clearly defined. It is an important advantage that differs from the existing techniques [133][134] that assume one of or both of the original signals is/are known.

## 8.2 MODELING OF NONLINEAR RECEIVER WITH TWO BAND-LIMITED INPUT SIGNALS

It is quite instructive to study the nonlinear mechanism of the RF front-end when two band-limited input signals are assumed, inasmuch as the methodology for nonlinear compensation is based firmly on the nonlinear model. We will return to the issues about compensator in due course, but now perform some simple analysis on the passband nonlinearity with the form of 5th-order power series:

$$\tilde{y}(t)=a_1\tilde{x}(t)+a_3\tilde{x}^3(t)+a_5\tilde{x}^5(t) \tag{8.1}$$

in which even-order terms are omitted since they do not generate spectral content near the frequency band of interest. The tildes are used to indicate the input and output RF signals $\tilde{x}(t)$ and $\tilde{y}(t)$ in passband. $a_i$ ($i$=1, 3, 5) is the parameter describing the nonlinear system.

The two signals are simultaneously received by the antenna so that the input of the LNA can be simply assumed as the addition of the two signals that can be expressed as

$$\tilde{x}(t) = \tilde{x}_L(t) + \tilde{x}_H(t), \tag{8.2}$$

where, as the one defined in (2.3)

$$\begin{aligned} \tilde{x}_L(t) &= A_L(t)\cos\left(\omega_L t + \phi_L(t)\right) \\ \tilde{x}_H(t) &= A_H(t)\cos\left(\omega_H t + \phi_H(t)\right) \end{aligned} \tag{8.3}$$

are the two concurrently received signals. By substituting (8.2) and (8.3) into (8.1) it yields the spectral contents at $\omega_L$ and $\omega_H$:

$$\begin{aligned} \tilde{y}_L(t) = \Big[ & a_1 A_L(t) + \frac{3}{4} a_3 A_L^3(t) + \frac{3}{2} a_3 A_L(t) A_H^2(t) + \frac{5}{8} a_5 A_L^5(t) + \\ & \frac{15}{4} a_5 A_L^3(t) A_H^2(t) + \frac{15}{8} a_5 A_L(t) A_H^4(t) \Big] \cos\left(\omega_L t + \phi_L(t)\right) \end{aligned} \tag{8.4}$$

and

$$\begin{aligned} \tilde{y}_H(t) = \Big[ & a_1 A_H(t) + \frac{3}{4} a_3 A_H^3(t) + \frac{3}{2} a_3 A_H(t) A_L^2(t) + \frac{5}{8} a_5 A_H^5(t) + \\ & \frac{15}{4} a_5 A_H^3(t) A_L^2(t) + \frac{15}{8} a_5 A_H(t) A_L^4(t) \Big] \cos\left(\omega_H t + \phi_H(t)\right). \end{aligned} \tag{8.5}$$

The second and third lines of (8.4) and (8.5) represent third and fifth "degree" of distortion products of most interest, in terms of their detrimental effects. We assume the equivalent base-band expressions for the two signals are given as:

$$\begin{aligned} x_L(t) &= A_L(t)\exp\left(j\phi_L(t)\right) \\ x_H(t) &= A_H(t)\exp\left(j\phi_H(t)\right), \end{aligned} \tag{8.6}$$

which recalling (2.2) are related to passband signal through the following equations:

$$\begin{aligned} \tilde{x}_L(t) &= \Re\left\{x_L(t)\exp(j\omega_L t)\right\} \\ \tilde{x}_H(t) &= \Re\left\{x_H(t)\exp(j\omega_H t)\right\}, \end{aligned} \tag{8.7}$$

By superseding (8.7) into (8.1), the respective base-band models at $\omega_L$ and $\omega_H$ can be obtained:

$$\begin{aligned} y_L(t) = {} & a_1 x_L(t) + \frac{3}{4} a_3 \left|x_L^2(t)\right| x_L(t) + \\ & \frac{3}{2} a_3 \left|x_H(t)\right|^2 x_L(t) + \frac{5}{8} a_5 \left|x_L(t)\right|^4 x_L(t) + \\ & \frac{15}{4} a_5 \left|x_H(t)\right|^2 \left|x_L(t)\right|^2 x_L(t) + \frac{15}{8} a_5 \left|x_H(t)\right|^4 x_L(t) \end{aligned} \tag{8.8}$$

and

$$\begin{aligned} y_H(t) = {} & a_1 x_H(t) + \frac{3}{4} a_3 \left| x_H^2(t) \right| x_H(t) + \\ & \frac{3}{2} a_3 \left| x_L(t) \right|^2 x_H(t) + \frac{5}{8} a_5 \left| x_H(t) \right|^4 x_H(t) + \\ & \frac{15}{4} a_5 \left| x_L(t) \right|^2 \left| x_H(t) \right|^2 x_H(t) + \frac{15}{8} a_5 \left| x_L(t) \right|^4 x_H(t). \end{aligned} \tag{8.9}$$

After absorbing the constant scales, they are simplified to

$$\begin{aligned} y_L(t) = {} & a_1' x_L(t) + a_3' \left| x_L^2(t) \right| x_L(t) + \\ & 2a_3' \left| x_H(t) \right|^2 x_L(t) + a_5' \left| x_L(t) \right|^4 x_L(t) + \\ & 6a_5' \left| x_H(t) \right|^2 \left| x_L(t) \right|^2 x_L(t) + 3a_5' \left| x_H(t) \right|^4 x_L(t) \end{aligned} \tag{8.10}$$

and

$$\begin{aligned} y_H(t) = {} & a_1' x_H(t) + a_3' \left| x_H^2(t) \right| x_H(t) + \\ & 2a_3' \left| x_L(t) \right|^2 x_H(t) + a_5' \left| x_H(t) \right|^4 x_H(t) + \\ & 6a_5' \left| x_L(t) \right|^2 \left| x_H(t) \right|^2 x_H(t) + 3a_5' \left| x_L(t) \right|^4 x_H(t). \end{aligned} \tag{8.11}$$

An important observation that can be made from inspection of (8.10) and (8.11) is the emergence of inter-band modulation terms, such as $|x_H(t)|^2 x_L(t)$, which have even higher amplitude than self-generated distortions such as $|x_L(t)|^2 x_L(t)$.

The analysis given before assumes two separated modulated carriers with fixed spacing. Notwithstanding not explicitly shown in the above derivation, it should be noted that there are also plenty of spectral content in other bands, such as in $2\omega_L+\omega_H$, $2\omega_H+\omega_L$, and so forth. The reason for omitting these terms is that they can be simply removed by band-pass filters in sequel and are not usable for the compensator. Another key point to note is that these two carriers have respective prescribed bandwidths so that the intermodulation bands stretch out to several times the band limits. For instance, $|x_H(t)|^2 x_L(t)$ produces a spectra with a bandwidth of $BW_L+2BW_H$, where $BW_L$ and $BW_H$ are respective bandwidth of $x_L(t)$ and $x_H(t)$. This is important since the spectral contents outside the prescribed band will be exploited by the compensator as the cost function, as will be presented hereinafter.

## 8.3 COMPENSATION OF THE NONLINEAR DISTORTION OF CONCURRENT DUAL-BAND RECEIVER

The proposed compensation scheme, referring to Fig.8.3, comprises two separated paths for capturing respective $x_L(t)$ and $x_H(t)$. The AD converters cover several times of the original signal bandwidths in order to retain the components stemming from self-generated distortion and inter-band intermodulation. The digitalized signals are then aligned with a special cross-

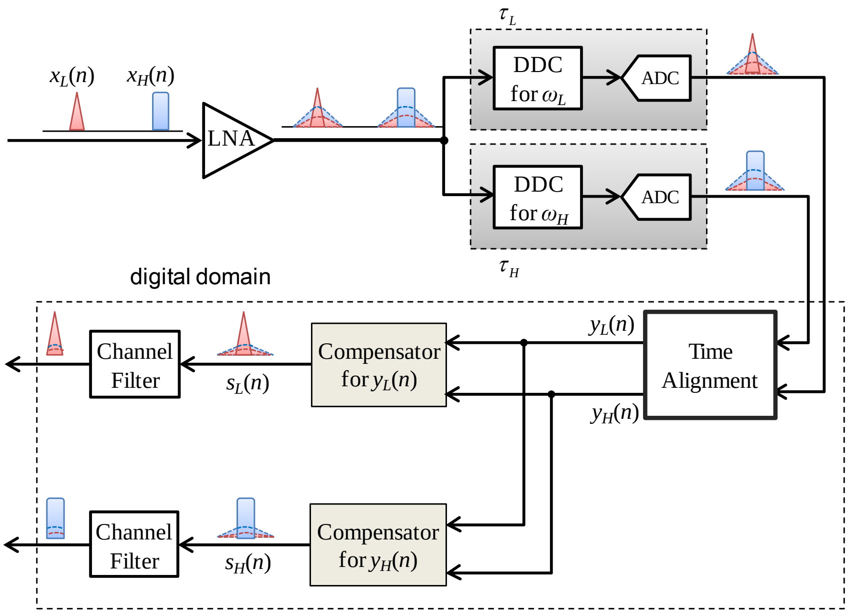


**Fig. 8.3 Block diagram of the linearization architecture of the concurrent dual-band receiver.**

correlation method as will be presented in Section 8.3.5. The aligned signals are jointly applied to estimate the parameters of the compensators. Having sketched the framework of the blind nonlinear compensation technique, the work left is to elaborate some important blocks involved in Fig.8.3.

## 8.3.1 The Inverse Model

From the discussions given in Section II, it is clear that the nonlinear models are bivariate polynomials of degree $P$ which can be simply denoted by

$$\begin{aligned} y_L(t) &= F_L\left(x_L(t), x_H(t)\right) \\ y_H(t) &= F_H\left(x_L(t), x_H(t)\right). \end{aligned} \tag{8.12}$$

Thus the inverse model can be represented by

$$\begin{aligned} s_L(t) &= G_L\left(y_L(t), y_H(t)\right) \\ s_H(t) &= G_H\left(y_L(t), y_H(t)\right), \end{aligned} \tag{8.13}$$

where $G_L$ and $G_H$ denote the inverse of $F_L$ and $F_H$, which are two bivariate polynomials of degree $Q$. In such case, the composite $F(G)$ response will have distortions up to $(QP)$-th order. Numerical optimization method will be used for estimation of $G_L$ and $G_H$.

Because of the symmetry of (8.10) and (8.11), only the compensator for $x_L(t)$ is considered, as the counterpart for $x_H(t)$ can be easily extrapolated. Similar to (8.10), the inverse bivariate polynomial model has the form of

$$\begin{aligned} s_L(t) = y_L(t) + c'_{30}\left|y_L(t)\right|^2 y_L(t) + \\ c'_{32}\left|y_H(t)\right|^2 y_L(t) + c'_{50}\left|y_L(t)\right|^4 y_L(t) + \\ c'_{52}\left|y_L(t)\right|^2 \left|y_H(t)\right|^2 y_L(t) + c'_{54}\left|y_H(t)\right|^4 y_L(t) + \cdots, \end{aligned} \tag{8.14}$$

note that the coefficient of the first term is set equal to 1, which is mandate as will be seen in the subsequent discussions. This model can be generalized to *Q*th-order as given in

$$s_L(t) = y_L(t) + \sum_{q=3,5\ldots}^{Q} \sum_{r=0,2,4\ldots}^{q-1} c'_{q,r}\left|y_L(t)\right|^{q-r-1}\left|y_H(t)\right|^{r} y_L(t), \tag{8.15}$$

where $c_{q,r}$ represents the respective coefficient. The diagram of the bivariate model up of *Q*th-order is shown in Fig.8.4.

### 8.3.2 Blind Nonlinear Compensation

Having developed the inverse model, the problem left is how to extract the parameters of the model. Since we have assumed the RF front-end is a black-box and there is no further knowledge of the received signals, the nonlinear compensation handled here represents a blind identification problem. Consider that a typical indication of nonlinear distortion is the occurrence of out-of-band emission, and thereby utilization of the out-of-band components becomes a natural consequence of nonlinear compensation [128][129]. The underlying principle hereof is to use a filter to obtain the distortion components excluding the bands of desired signals. Then, the out-of-band emission is exploited as the cost function, while the adaption aims to minimization of this emission.

For notional simplicity, the compensator (inverse model) given in (8.15) can be expressed as

$$s_L(t) = y_L(t) + \mathbf{v}(t)\mathbf{C}' \tag{8.16}$$

where

$$\begin{aligned} \mathbf{v}(t) = \Big[ \left|y_L(t)\right|^2 y_L(t), \left|y_H(t)\right|^2 y_L(t), \\ \left|y_L(t)\right|^4 y_L(t), \left|y_H(t)\right|^2 \left|y_L(t)\right|^2 y_L(t), \cdots, \\ \left|y_H(t)\right|^{Q-3} \left|y_L(t)\right|^2 y_L(t), \left|y_H(t)\right|^{Q-1} y_L(t) \Big] \end{aligned} \tag{8.17}$$

is the data matrix of the nonlinear functions contained in the polynomial, and $\mathbf{C}'=[c'_{30}, c'_{32}, c'_{50}, \ldots, c'_{QQ-1}]^{\mathrm{T}}$ is the coefficient vector for the corresponding terms. Note that there exists an implicit assumption that $c'_{10}=1$, which indicates that the linear term $y_L(t)$ has a gain of 1. The scaling that $c'_{10}\neq 1$ is also applicable since the constant scaling can be easily restored. The reason for fixing $c_{10}$ is to prevent the signal from extinction, as one set of coefficients for minimizing the out-of-band component is to set all of the coefficients (including $c_{10}$) to

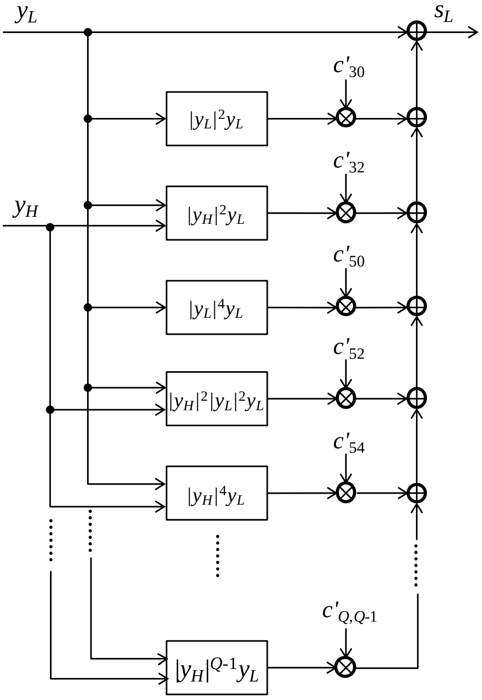


**Fig. 8.4 Diagram of the baseband bivariate polynomial model.**

0.

Since the objective of the nonlinear compensation is to eliminate the out-of-band spectral contents, see Fig.8.5, the cost function can be defined as

$$J(\mathbf{C}') = \left\|\hat{s}_L(t)\right\|_2^2 = \left\|\hat{y}_L(t) + \hat{\mathbf{v}}(t)\mathbf{C}'\right\|_2^2, \tag{8.18}$$

where the hat denotes the data outside the specified bands. The respective signals are obtained by applying a high-pass filter whose impulse response is $h_L(t)$:

$$\begin{aligned} \hat{\mathbf{v}}(t) &= \mathbf{v}(t) * h_L(t) \\ \hat{y}_L(t) &= y_L(t) * h_L(t) \\ \hat{s}_L(t) &= s_L(t) * h_L(t), \end{aligned} \tag{8.19}$$

where * denotes circular convolution to make the filter output length comply with the input length.

It has been mentioned above that the model fixes the linear term to prevent the signal from extinction, which, whereas, creates an obvious bottleneck on the performance, in respect that the linear term is correlated with the remaining terms in the model. Directly implementing the model in (8.15) may not achieve satisfactory performance and even in some cases become unworkable. This problem can be solved by de-embedding the linear term from the nonlinear terms, which is similar to last chapter and will be shown in the subsequent section.

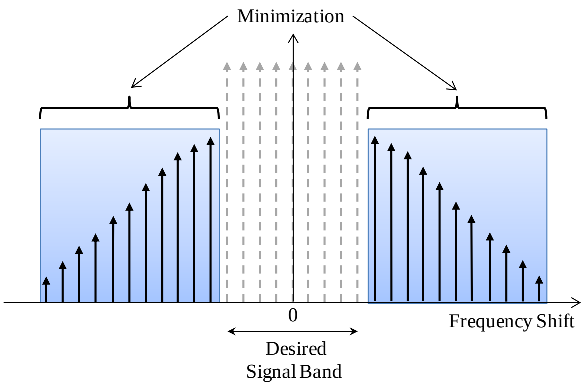


**Fig. 8.5 Minimization object, which is composed on spectral contents outside the desired signal band.**

### 8.3.3 De-embedding the Linear Term

The nonlinear terms in a polynomial shown in (8.15) have high correlation with the linear term $y_L(t)$, and this will affect the parameter extraction as the linear term is fixed. Consequently, it is reasonable to de-embed the linear term from the nonlinear terms, so that the cost function can be uncorrelated with the linear term and thereby the performance of the compensator can be enhanced. De-embedding is accomplished by simply subtracting the linear term from the nonlinear terms, which yields the model diagrammatically depicted in Fig.8.6 and has the form of

$$s_L(t) = y_L(t) + \mathbf{\Phi}(t)\mathbf{C} \tag{8.20}$$

with

$$\mathbf{\Phi}(t) = [\varphi_{30}(t), \varphi_{32}(t), \varphi_{50}(t), \cdots, \varphi_{P0}(t), \cdots, \varphi_{P(P-1)}(t)], \tag{8.21}$$

where

$$\varphi_{qr}(n) = \left|y_L(t)\right|^{q-r-1} \left|y_H(t)\right|^r y_L(t) - \beta_{qr} y_L(n) \tag{8.22}$$

and $\beta_{qr}$ are determined parameters according to the probability distribution function of the signal. With this structure, the nonlinear terms are explicitly orthogonal with the linear term.

Having developed the model in (8.20), the question left lies on the values of parameter $\mu_{qr}$, which can be determined by the following equation

$$E\left\{\left(\left|y_L(t)\right|^{q-r-1} \left|y_H(t)\right|^r y_L(t) - \beta_{qr} y_L(t)\right) y_L^*(t)\right\} = 0, \tag{8.23}$$

where E{· } is the expectation and the superscript * denotes the conjugation. Finally, (8.23) yields

$$\beta_{qr} = \frac{E\{\left|y_L(t)\right|^{q-r+1} \left|y_H(t)\right|^r\}}{E\{\left|y_L(t)\right|^2\}}. \tag{8.24}$$

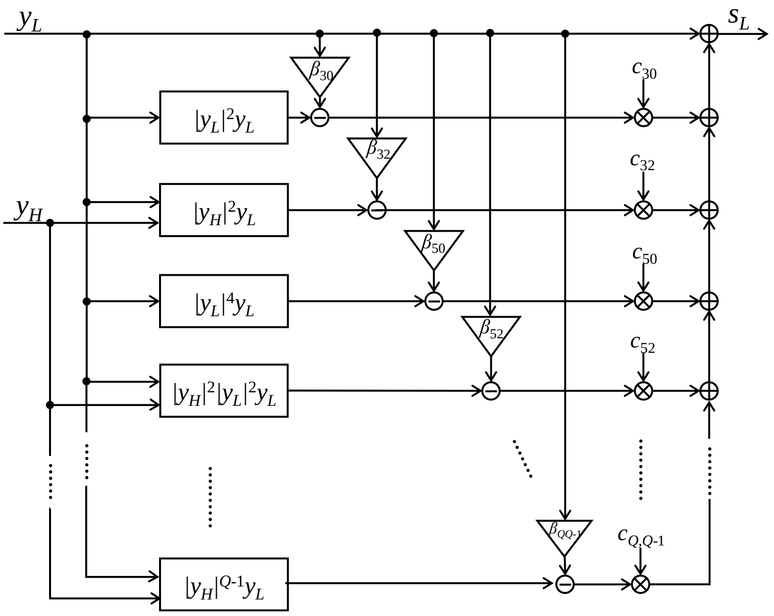


**Fig. 8.6 Diagram of the compensator model with linear term de-embedding.**

## 8.3.4 Iterative Algorithm for Blind Nonlinear Compensation

The well-established least squares (LS) algorithm is most common for system identification. As a consequence of the requirement for use of a long sequence of the observed signal, LS is computational ineffective. On the other hand, sample-by-sample-based methods, such as least mean squares (LMS) and recursive least squares (RLS), are less robust, and the updating for each individual sample is not necessary for most applications. In such sense, the block-based iteration method proposed in this section provides a good balance between the complexity and adaption speed.

Using vector notation, the signal for $k$-th iteration can be expresses as

$$\mathbf{y}_L^{(k)} = \begin{bmatrix} y_L\left((k-1)\mathrm{L}+1\right) \\ y_L\left((k-1)\mathrm{L}+2\right) \\ \vdots \\ y_L\left(k\mathrm{L}\right) \end{bmatrix}, \tag{8.25}$$

and similarly, $\mathbf{\Phi}^{(k)}$ can be defined as

$$\mathbf{\Phi}^{(k)} = [\boldsymbol{\varphi}_{30}^{(k)}, \boldsymbol{\varphi}_{32}^{(k)}, \boldsymbol{\varphi}_{50}^{(k)}, \cdots, \boldsymbol{\varphi}_{PP-1}^{(k)}], \tag{8.26}$$

where

$$\boldsymbol{\varphi}_{qr}^{(k)} = \left|\mathbf{y}_L^{(k)}\right|^{q-r-1} \left|\mathbf{y}_H^{(k)}\right|^{r} \mathbf{y}_L^{(k)} - \beta_{qr}^{(k)} \mathbf{y}_L^{(k)}, \tag{8.27}$$

therein $\beta_{qr}^{(k)}$ can be determined by replacing the expectation operator in (8.24) by sample average

$$\beta_{qr}^{(k)} = \frac{\left(\mathbf{y}_L^{(k)}\right)^H \left(\left|\mathbf{y}_L^{(k)}\right|^{q-r-1} \left|\mathbf{y}_H^{(k)}\right|^r \mathbf{y}_L^{(k)}\right)}{\left(\mathbf{y}_L^{(k)}\right)^H \mathbf{y}_L^{(k)}}. \tag{8.28}$$

The compensator produces the output

$$\mathbf{s}_L^{(k)} = \mathbf{y}_L^{(k)} + \mathbf{\Phi}^{(k)}\mathbf{C}^{(k)}, \tag{8.29}$$

where $\mathbf{C}^{(k)}$ is the parameter vector of the compensator for *k*-th iteration, and which can be set to all zeros initially. Then a high-pass filter is applied to obtain respective signals out of the prescribed band, and it yields

$$\begin{aligned} \hat{\mathbf{u}}^{(k)} &= \mathbf{H}\mathbf{u}^{(k)} \\ \hat{\mathbf{y}}_L^{(k)} &= \mathbf{H}\mathbf{y}_L^{(k)} \\ \hat{\mathbf{s}}_L^{(k)} &= \mathbf{H}\mathbf{s}_L^{(k)}, \end{aligned} \tag{8.30}$$

where **H** is a Toeplitz matrix containing the impulse response of the high-pass filter $h_L(t)$ and it can be defined as

$$\mathbf{H} = \begin{bmatrix} h_L(1) & h_L(2) & \cdots & h_L(N) \\ h_L(N) & h_L(1) & \cdots & h_L(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ h_L(2) & h_L(3) & \cdots & h_L(1) \end{bmatrix}. \tag{8.31}$$

The cost function can be written as

$$\begin{aligned} J(\mathbf{C}^{(k)}) &= \left\|\hat{\mathbf{s}}_L^{(k)}\right\|_2^2 \\ &= \left\|\hat{\mathbf{y}}_L^{(k)} + \hat{\mathbf{u}}^{(k)}\mathbf{C}^{(k)}\right\|_2^2 \\ &= \left\|\mathbf{H}(\mathbf{y}_L^{(k)} + \mathbf{u}^{(k)}\mathbf{C}^{(k)})\right\|_2^2. \end{aligned} \tag{8.32}$$

The iterative method exploiting damped Gauss-Newton (DGN) iteration has the updating function for the *k*-th iteration of

$$\mathbf{C}^{(k+1)} = \mathbf{C}^{(k)} - \varepsilon \frac{J'(\mathbf{C}^{(k)})}{2(\hat{\mathbf{u}}^{(k)})^H \hat{\mathbf{u}}^{(k)}} \tag{8.33}$$

where

$$\begin{aligned} J'(\mathbf{C}^{(k)}) &= \frac{\partial J(\mathbf{C}^{(k)})}{\partial\left(\mathbf{C}^{(k)}\right)^H} \\ &= 2\left((\hat{\mathbf{u}}^{(k)})^H \hat{\mathbf{y}}_L^{(k)} + (\hat{\mathbf{u}}^{(k)})^H \hat{\mathbf{u}}^{(k)}\mathbf{C}^{(k)}\right) \\ &= 2(\hat{\mathbf{u}}^{(k)})^H \hat{\mathbf{s}}_L^{(k)} \\ &= 2(\mathbf{u}^{(k)})^H \mathbf{H}^H \mathbf{H}\mathbf{s}_L^{(k)}. \end{aligned} \tag{8.34}$$

By substituting (8.33) with $J'(\mathbf{C}^{(k)})$ in (8.34) there yields

$$\mathbf{C}^{(k+1)} = \mathbf{C}^{(k)} - \varepsilon\left((\mathbf{u}^{(k)})^H \mathbf{H}^H \mathbf{H}\mathbf{u}^{(k)}\right)^{-1} (\mathbf{u}^{(k)})^H \mathbf{H}^H \mathbf{H}\mathbf{s}_L^{(k)} \tag{8.35}$$

where $\varepsilon$ is the step size.

With block-by-block operation, DGN iteration maximizes the efficiency of the system by distributing the fixed process overhead across many samples. In addition, DGN absorbs the merits of LS, such as high robustness, and enables adaptive update of the models. Also, the length of the vectors $N$ can be much smaller than the ones in LS, hence the required memory space and computation load are much lower. Another important merit of this algorithm is that the value of the $\mu$, which is determined by the signal probability distribution, can be updated in each iteration step. As a consequence, the non-correlation condition set in (8.23) can be strictly met.

## 8.3.5 Time Alignment

$x_L(t)$ and $x_H(t)$ experience different delay paths transforming from RF to base-band, and the delay values are denoted as $\tau_d$ and $\tau_i$, respectively, in Fig.8.3. Fortunately, these two values are constant as long as the circuits keep unchanged. An offline calibration, as a consequence, can be applied to determine these delays. However, in order to fully exploit the advantages of digital signal processing and abridge the calibration procedure, an automatic alignment method is considered in this section.

A commonly used method for delay estimation is to correlate the two signals, and find the delay value that maximizes or minimizes the correlation function. Such a method is actually an exhaustive attempt of all possible delays in the prescribed range. Indeed, there is a prerequisite that the two signals should have some mutual *correlation* to ensure the existence of an obvious peak of the cross correlation function. Unfortunately, nevertheless, $x_L(t)$ and $x_H(t)$ are completely independent, and thus directly calculating the cross correlation of $y_L(t)$ and $y_H(t)$ will not produce an obvious peak, owing the fact that $y_L(t)$ and $y_H(t)$ have very weak correlation with each other.

By inspecting (8.10) and (8.11), some inter-band terms that represent the weak *correlation* between the two signals can be found. These terms, scaled by $a_3$ and $a_5$, have much lower power, i.e. several tens in decibel, lower than the linear term with a gain of $a_1$. Therefore, exempting the linear component becomes a nature consequence of amplifying the *correlation*. This can be accomplished by filtering out the in-band spectral contents and leaving the out-of-band distortion components. By assuming $h_L(t)$ and $h_H(t)$ two high-pass filters with stop bands coherent with the bandwidth of $x_L(t)$ and $x_H(t)$, the delay estimation can be expressed as

$$\begin{aligned}\Delta\tau &= \tau_L - \tau_H \\ &= \arg_\tau \max\left\{xcorr\left[|o_L(t)|, |o_H(t)|\right]\right\} \\ &= \arg_\tau \max\left\{\sum_{t=0}^{T} |o_L(t)||o_H(t-\tau)|\right\},\end{aligned} \tag{8.36}$$

where

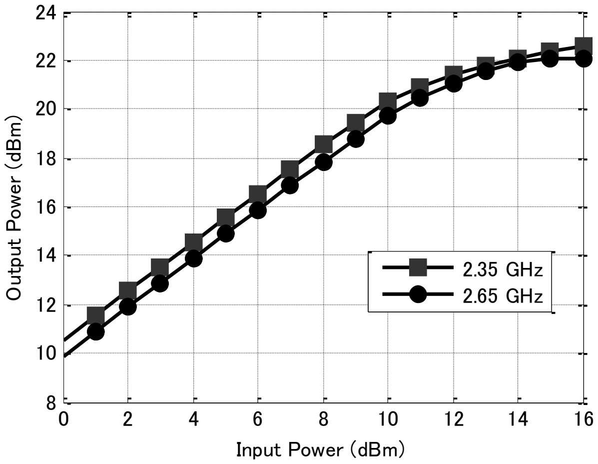


**Fig. 8.7 Measured swept power data of the amplifier in test which were obtained in 2.35 GHz and 2.65 GHz.**

$$\begin{aligned} o_L(t) &= y_L(t) * h_L(t) \\ o_H(t) &= y_H(t) * h_H(t) \end{aligned} \tag{8.36}$$

and $T$ determines the range of searching.

## 8.4 EXPERIMENTAL RESULTS

Having introduced the respective approaches above, the tools now are available to build a concurrent dual-band receiver with adaptive nonlinearity compensation. In this section, experiments were carried out to evaluate the suitability of the proposed method. A LNA, which is HMC374 [136] from Hittite, providing a gain of over 10 dB was used in the test. As the LNA has a wide and flat operation frequency range (0.3 -3 GHz), its memory effect can be neglected. The nonlinear nature of the amplifier is apparent in the results of continuous wave (CW) tests shown in Fig.8.7, from which it can be seen that the input saturation

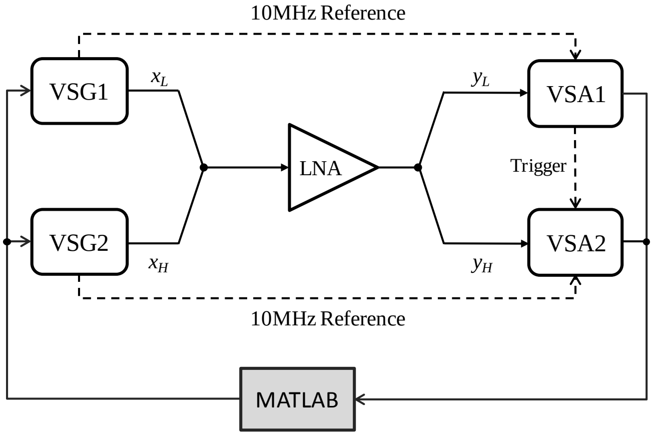


**Fig. 8.8 Outline of the measurement setup.**

point is around 16 dBm and the 1dB compression point around 11 dBm. Two vector signal generators (VSGs), both of which are Agilent N5192a, were applied to generate the signals in different bands. Vector signal analyzers (VSAs) with 110MHz acquisition bandwidth, which are RSA6100a from Tektronix, were employed to digitalize and transfer the signals to base-band. The base-band signals were processed in MATLAB and the whole measurement setup of the dual-band receiver linearization is presented in Fig.8.8.

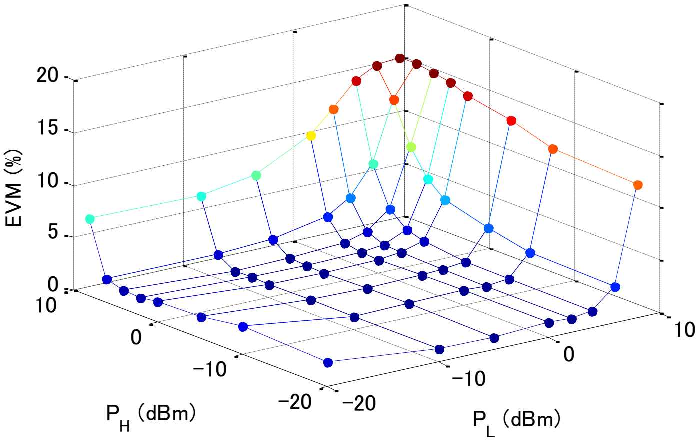
In the actual operation, the power of the signal arriving at the receiver has a large dynamic range depending on the condition of the channel, therein automatic gain control (AGC) is required to extend the receiver dynamic range. In our test, the attenuator of the VSA was adjusted to imitate the function of AGC.

The measurement applies the following signals and parameters:

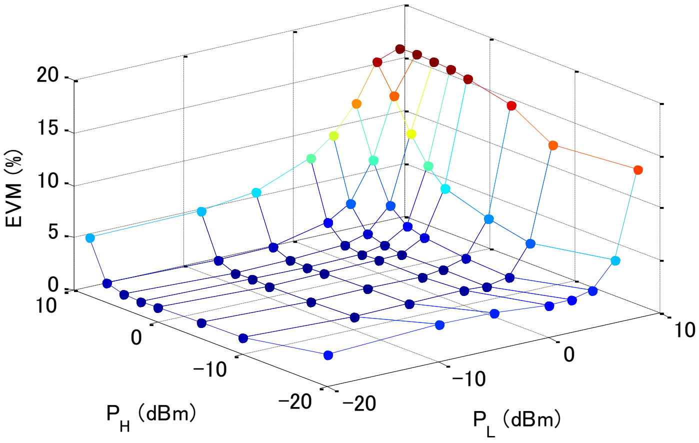
- $x_L(t)$ is an LTE signal: 64-QAM for all subcarriers, 2350 MHz center frequency, 10 MHz bandwidth and PAPR (@99.999% probability) of 10.13 dB.
- $x_H(t)$ is also an LTE signal: 64-QAM for all subcarriers, 2650 MHz center frequency, 20 MHz bandwidth and PAPR (@99.999% probability)of 10.65 dB.
- The block size $N$ is set to 512, and the step size $\beta$ for iteration is set to 1 and iteration time to 5.

The computer used in the measurement contains an Intel I3 CPU and 4 GB memory. The execution time in MATLAB for one time of iteration is around 3.2 ms and the total execution time (including 5 times of iterations for the two respective signals) is around 35 ms. To access the performance of the whole system, error vector magnitude (EVM) was chosen as the figure of merit. The measured EVMs for both $y_L(t)$ and $y_H(t)$ are illustrated in Fig.8.9 to provide an intuitive view over the receiver performance without compensator. The EVMs were obtained by exhaustive measurements under different average power of $x_L(t)$ and $x_H(t)$ say, $P_L$ and $P_H$. The EVM performance, as can be inspected from Fig.8.9, is dependent upon with the power of both $x_L(t)$ and $x_H(t)$. As $P_L$ or/and $P_H$ getting stronger, the EVMs for both of the signals degrade as a result of decreased signal-to-distortion ratio (SDR). The EVM also degrades as the respective signal power reduces below -15 dBm taking for granted the reduced signal to noise radio (SNR). Note that as the power of one signal is much lower than the other one, the power of the strong signal will have a deterministic effect on the EVMs for both of the two signals. This case tends to happen in a circumstance that the receiver is close to one base station while being far from the other. Since we only concern the nonlinear distortion of the receiver, the subsequent measurements only focus on the region of high power and omit the influence of SNR reduction (region of low power).

Fig.8.10 shows the spectral plots of $y_L(t)$ and $y_H(t)$ with compensators of 3rd, 5th and 7th orders. As can be expected, the compensator of higher order yields better performance. In the experiments followed, the compensator order was set to 7, in order to achieve satisfactory performance. The spectral plots of $y_L(t)$ and $y_H(t)$ both with and without compensation are depicted in left hand of Fig.8.11 (a) and (b), where the power of $x_L(t)$ and $x_H(t)$ was -10 dBm and 6 dBm, respectively. Note that the backoff of $x_H(t)$ here was comparable with the PAPR of the LTE signal, which is about 10 dB, so that a small proportion of $x_H(t)$ entered the saturation region, and this represents a severe distortion for 64-QAM signal. Consider that the power of $x_L(t)$ was set much lower than $x_H(t)$, the spectral regrowth was mainly caused by inter-band modulation. With compensation, the undesired spectral contents decrease by



(a) Measured EVMs for $y_L(t)$

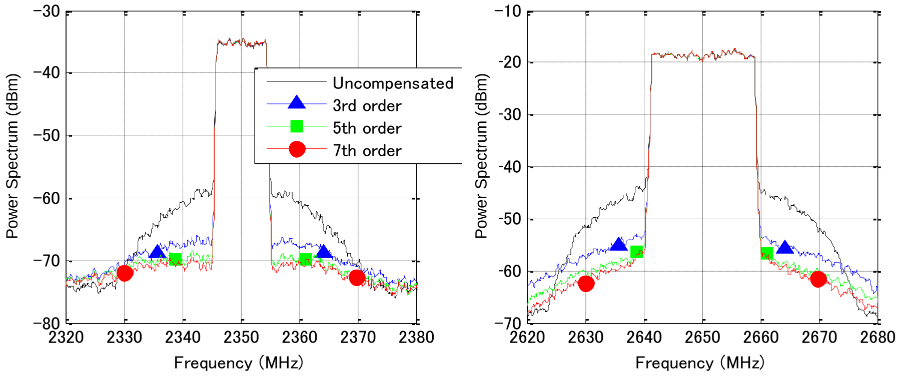


(b) Measured EVMs for $y_H(t)$

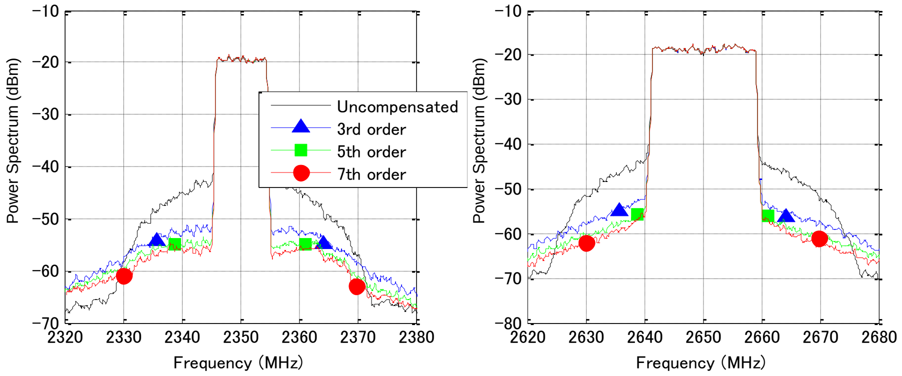
**Fig. 8.9 The measured EVMs for $y_L(t)$ and $y_H(t)$ by sweeping respective input power $P_L$ and $P_H$.**

around 15 dB for both $y_L(t)$ and $y_H(t)$. In addition, the merit of the proposed compensator can also be visualized by the constellation plots shown in right hand of Fig.8.11. The measured results shown in Fig.8.12, on the other hand, were obtained by setting $P_L$ and $P_H$ both to 6 dBm. In this case, substantial self-generated distortion and inter-band modulation contents were generated, as can be observed from Fig.8.12. It is worth to mention that the case in Fig.8.12 may be avoided by lowering the gain with AGC in practice. However, decreasing the gain to avoid nonlinear distortion for the case in Fig.8.11 would result in loss of sensitivity of the receiver for weak signal. In order to sustain enough sensitivity, the nonlinear distortion that occurs in this case tends to be unavoidable; hence the compensator becomes essential for achieving higher sensitivity and linearity.

The EVMs obtained by sweeping $P_H$ are plotted in Fig.8.13, from which we can see that



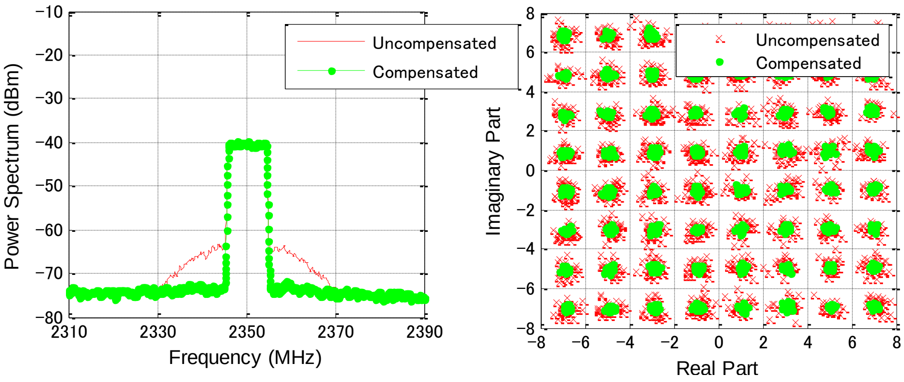
(a) The powers of $x_L(t)$ and $x_H(t)$ are -10 dBm and 6 dBm, respectively.



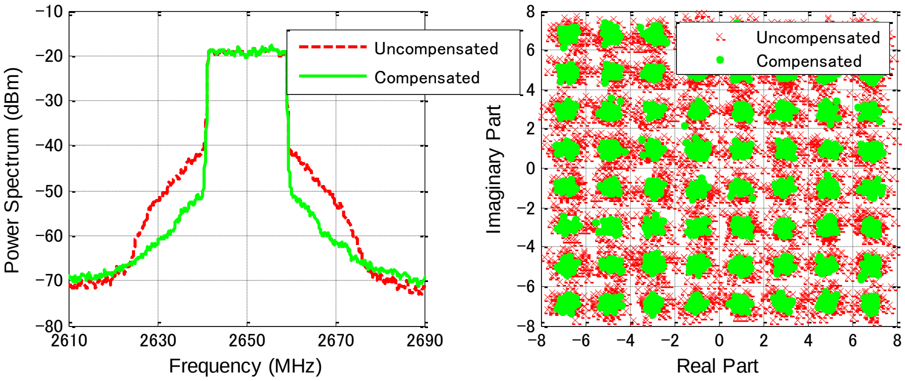
(b) The powers of $x_L(t)$ and $x_H(t)$ are both 6 dBm.

**Fig. 8.10 The spectral plots for $y_L(t)$ and $y_H(t)$ with compensators of different orders.**

$x_H(t)$ has significant impact on the $x_L(t)$ as its power exceeds 5 dBm. Furthermore, as $P_H$ approaches 10 dBm, $x_H(t)$ is severely saturated accompanying with serious EVM degradation. Additionally, with the power of $x_L(t)$ getting stronger, for instance $P_L$=7 dBm, the EVM degrades significantly for $y_H(t)$, see Fig.8.13(b). On the other hand, with the proposed compensator, obvious EVM improvement for both $y_L(t)$ and $y_H(t)$ can be achieved, especially when the power of both the signals is high.

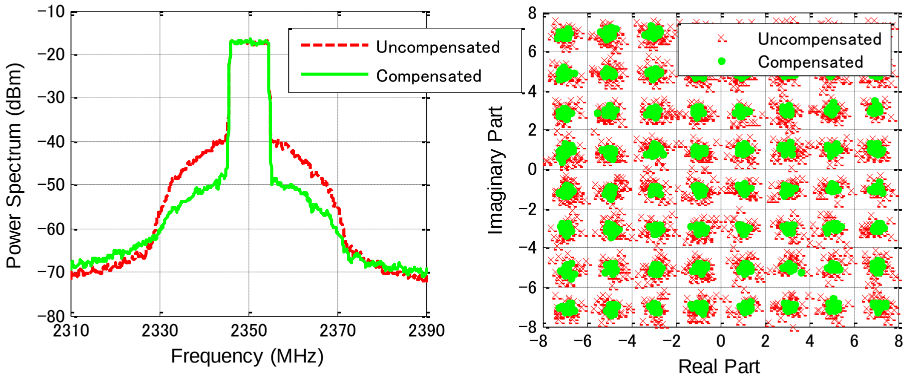


(a) Spectral plots and the constellations for $y_L(t)$

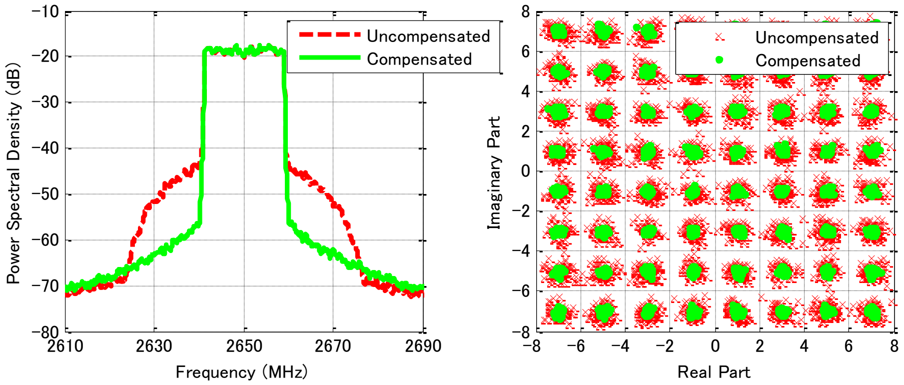


(b) Spectral plots and the constellations for $y_H(t)$

**Fig. 8.11 The spectral plots and the constellations for $y_L(t)$ and $y_H(t)$ with and without the compensator, where the power of $x_L(t)$ is -10 dBm and the power of $x_H(t)$ is 6 dBm.**

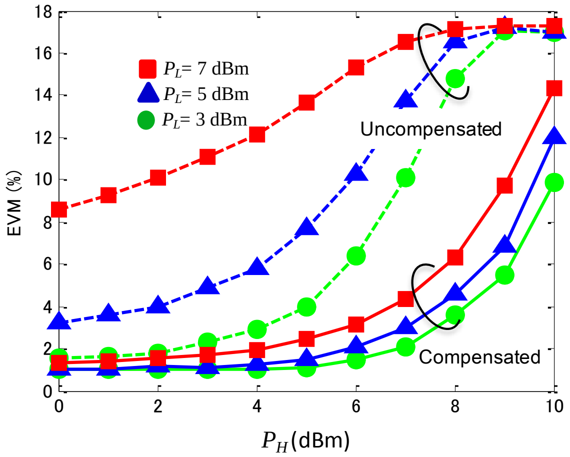


(a) Spectral plots and the constellations for $y_L(t)$

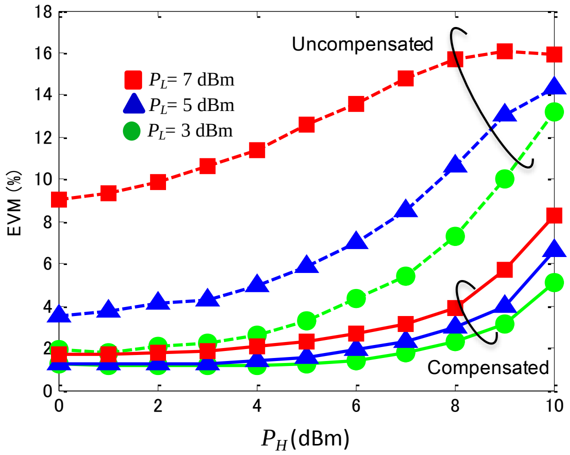


(b) Spectral plots and the constellations for $y_H(t)$

**Fig. 8.12** **The spectral plots and the constellations for $y_L(t)$ and $y_H(t)$ with and without the compensator, where the power of $x_L(t)$ and $x_H(t)$ was both set to 6dBm.**



(a) Measured EVM for $x_L(t)$



(b) Measured EVM for $x_H(t)$

**Fig. 8.13 Measured EVMs with the power of $x_H(t)$ swept, where the powers of $x_L(t)$ were set to 3, 5 and 7 dBm, respectively.**

# Chapter 9.

# Spectra-Folding Feedback Architecture for Concurrent Dual-Band Power Amplifier Predistortion

This chapter proposes a spectra-folding feedback (SFFB) architecture for digital predistortion (DPD) of concurrent dual-band power amplifiers (PAs). Conventional dual-band DPD contains a dual-branch receiving path to feedback the PA output signals at respective bands. The proposed architecture, however, is constructed by a single-branch receiving path, so that the hardware expense (such as down converters, filters, analog-to-digital converters, etc.) of the feedback path is halved. In this architecture, simple modifications are made to the down converter while the operation of the ADC is kept unchanged. With proposed signal processing in the baseband, the SFFB DPD can achieve comparative performance as the conventional dual-branch DPD. Experiments were performed using various configurations of LTE/LTE-Advanced (long term evolution/ long term evolution advanced) signals of up to 80 MHz contiguous bandwidth and carrier spacing of 1 GHz. It is clearly shown by the experimental results that the proposed SFFB DPD can achieve satisfactory linearization performance, with significantly reduced hardware complexity.

## 9.1 CONCURRENT DUAL-BAND TRANSMISSION

The surge of applications of multi-standard and carrier aggregation [49]-[51] in wireless communication systems forces the use of multiple spectra that locate in different bands. This trend, however, requires the transmitter to embrace more parts of radio frequency (RF) systems. In addition, in order to be competitive, commercial wireless transmitter systems have to be efficient in terms of power, to be compliant with multiple standards, consume less energy, be more environmentally friendly, be of smaller form factor and, last but not least, produced with low manufacturing costs. Motivated by these aggressive designing goals, such it calls for circuit and structure breakthroughs.

Enabled by advanced power amplifier (PA) technologies, the concurrent dual-band transmitter, which can effectively reduce the bill of materials (BOM) cost paid for RF front-end by using a single PA, has become quite popular in recent years [137]-[138]. However, the distortion caused by the basic nonlinear characteristics of the PA causes an urgent need for the study on digital predistorter (DPD), which is a widely used tool for conventional PA

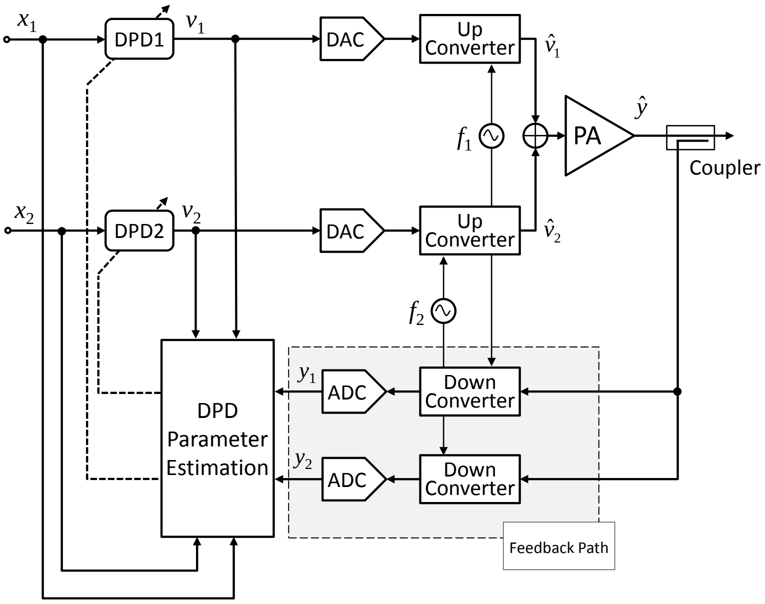


**Fig. 9.1 Illustration of the basic structure of concurrent dual-band transmitter with predistortion.**

linearization.

The dual-band predistorter, in contrast to its conventional counterpart, handles the two signals separately in baseband and transfers them to RF signals with two independent transmission paths [139], as shown in Fig.9.1. To obtain the distorted signals in respective frequency bands, two branches are necessary to convert the signals to baseband for DPD parameter estimation. It is clear from the figure that the dual-band transmission, compared with the traditional schemes, yields lower cost by using a single PA, but does not give any reduction for the BOM of DPD. The feedback paths are extra costs, which actually are only necessities for DPD rather than the transmitter. Especially, as the analog-to-digital converters (ADCs) tend to be the most expensive components and critical power consumers in the digital parts [21], it follows that designing simpler feedback paths can yield significant savings for dual-band DPD implementation.

The authors in [140] the first time presented the possibility to use an under-sampling ADC in a single-branch feedback path to receive the two signals simultaneously. The basic idea is similar to the so-called harmonic sampling [141], which converts the signals in different bands to the first Nyquist zone by smartly selecting the sampling frequency. This method was further discussed in details in [142], which also removed the use of down converters. Nevertheless, the experimental results shown in [142] were obtained from an ideal measurement, in which the signals were sampled with a high-speed ADC and subsampling was carried out by post-processing. It is worth to note that most off-the-shelf ADC products have finite analog bandwidth, which limits their application in subsampling mode.

In this chapter, a spectra-folding feedback (SFFB) architecture for dual-band PA predistortion is proposed. The corresponding identification method for DPD parameter estimation is also introduced. In this architecture, simple modifications are made to the down converter to yield a spectra folded intermediate frequency (IF) signal while the operation of the ADC is kept unchanged. The BOM cost of the feedback path is roughly half of the conventional dual-band predistortion, with no sacrifice in linearization performance. Although the proposed architecture may not be seen as very groundbreaking, we strongly believe that it has the potential to replace the existing dual-branch architecture and is rather promising for practical dual-band predistortion implementations.

This chapter is organized as follows: to better understand the concept and appreciate the advantage of the proposed architecture, we will briefly review the principle of the conventional dual-band predistorter in Section 9.2 first, and return to the due course in Section 9.3, in which the working principle and practical implementation issues of the SFFB architecture are presented. Experimental results are shown in Section 9.4 to validate the feasibility of the proposed architecture. Finally, a brief conclusion is drawn in Section 9.5.

## 9.2 DIGITAL PREDISTORTION OF CONCURRENT DUAL-BAND TRANSMITTER

In Fig.9.1 we use $x_1$ and $x_2$ to denote the two baseband signals to be transmitted to lower side band (LSB) and upper side band (USB), respectively, and the corresponding RF signals can be written as:

$$\begin{aligned}\tilde{x}_1(t) &= |x_1(t)|\cos\left(2\pi f_1 t + \angle x_1(t)\right)\\ \tilde{x}_2(t) &= |x_2(t)|\cos\left(2\pi f_2 t + \angle x_2(t)\right),\end{aligned} \tag{9.1}$$

where the tilde ~ is used to denote the signals in pass-band as defined in Chapter 2. As in last chapter, the nonlinear behavior for dual-band PA is represented with a polynomial, i.e. a power series with $Q$th-order:

$$\begin{aligned}\tilde{y}(t) &= \sum_{q=1}^{Q} \hat{a}_q \tilde{x}^q(t)\\ &= \sum_{q=1}^{Q} \hat{a}_q \left[\tilde{x}_1(t) + \tilde{x}_2(t)\right]^q,\end{aligned} \tag{9.2}$$

By substituting (9.1) into (9.2) and expanding the equation, it is easy to derive the output signals in prescribed bands. We will not proceed, as this procedure is very similar to a two-tone analysis that can be found in many textbooks. As it is assumed that there are proper matching networks or filters after the PA, we can only care about the signals that are located at $f_1$ and $f_2$, which are transferred to baseband with two feedback paths, see Fig.9.1. The equivalent baseband model for dual-band PA can be obtained:

$$\begin{aligned}\tilde{y}(t) &= \sum_{q=1}^{Q} \hat{a}_q \tilde{x}^q(t) \\ &= \sum_{q=1}^{Q} \hat{a}_q \left[\tilde{x}_1(t) + \tilde{x}_2(t)\right]^q,\end{aligned} \tag{9.3}$$

This expression can also be extended to memory model by introducing some delay taps. In addition, some even-order terms can be introduced into (9.3) to form a more general model. Here, we use the following model [135], which in discrete time has the form of

$$\begin{aligned}y_1(n) &= \sum_{m=0}^{M-1}\sum_{q=1}^{Q}\sum_{r=0}^{q-1} a_{q,r,m} \left|x_1(n-m)\right|^{q-r-1} \left|x_2(n-m)\right|^{r} x_1(n-m) \\ y_2(n) &= \sum_{m=0}^{M-1}\sum_{q=1}^{Q}\sum_{r=0}^{q-1} b_{q,r,m} \left|x_2(n-m)\right|^{q-r-1} \left|x_1(n-m)\right|^{r} x_2(n-m).\end{aligned} \tag{9.4}$$

The above model is linear regressive thereby we can give the model an algebraic structure through the kernel function defined below

$$\begin{aligned}K(x_1, x_2) = \Big[ x_1(\mathrm{n}) \ ,\cdots, &\left|x_1(n-m)\right|^{q-r-1} \left|x_2(n-m)\right|^{r} x_1(n-m), \cdots, \\ &\left|x_2(n-M-1)\right|^{Q-1} x_1(n-M-1) \Big].\end{aligned} \tag{9.5}$$

It is easy to see that the DPD outputs can be written as

$$\begin{aligned}v_1 &= K(\mathbf{x}_1, \mathbf{x}_2)\mathbf{W}_1 \\ v_2 &= K(\mathbf{x}_2, \mathbf{x}_1)\mathbf{W}_2,\end{aligned} \tag{9.6}$$

where $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{C}^{\frac{MQ(Q+1)}{2}}$ are the coefficient vectors for respective DPDs. The least squares (LS) is commonly used to extract the DPD coefficients,

$$\begin{aligned}\mathbf{W}_1 &= \left[K^H(\mathbf{y}_1, \mathbf{y}_2) K\ (\mathbf{y}_1, \mathbf{y}_2)\right]^{-1} K^H(\mathbf{y}_1, \mathbf{y}_2)\mathbf{v}_1 \\ \mathbf{W}_2 &= \left[K^H(\mathbf{y}_2, \mathbf{y}_1) K\ (\mathbf{y}_2, \mathbf{y}_1)\right]^{-1} K^H(\mathbf{y}_2, \mathbf{y}_1)\mathbf{v}_2,\end{aligned} \tag{9.7}$$

where $\mathbf{v}_1, \mathbf{v}_2, \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{C}^L$ are signal vectors and $L$ is the length.

## 9.3 SPECTRA-FOLDING FEEDBACK ARCHITECTURE FOR CONCURRENT DUAL-BAND PREDISTORTION

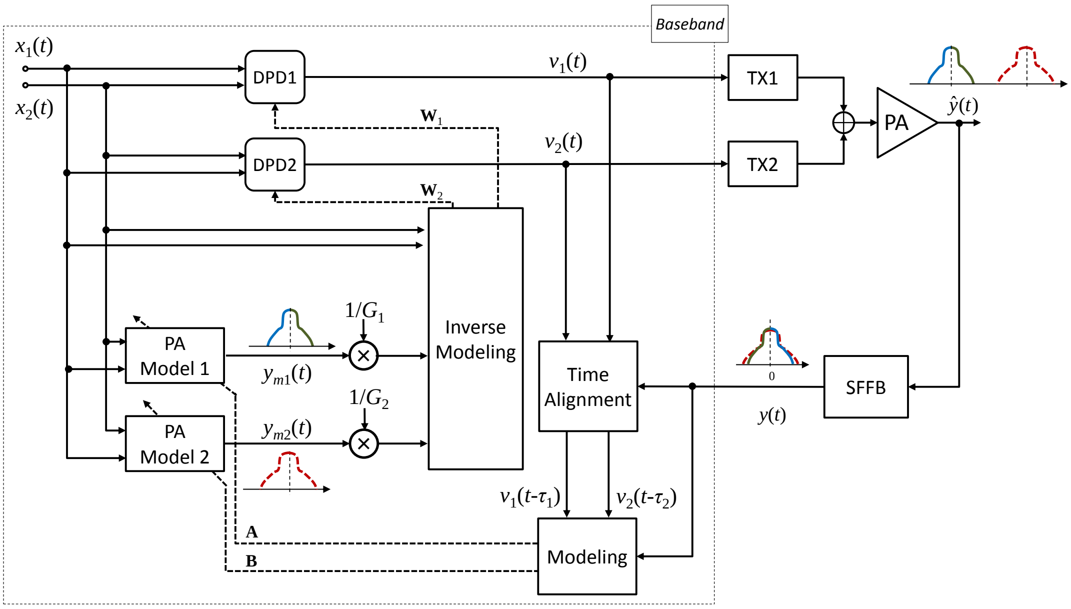


**Fig. 9.2 Illustration of the proposed SFFB architecture for concurrent dual-band predistorter.**

## 9.3.1 Basic Operation of Spectra-folding feedback Architecture

The proposed architecture is plotted in Fig.9.2, with the single feedback path being highlighted. Here, only one mixer is used to downconvert the two signals to IF and a low-pass filter (LPF) is employed to eliminate undesired frequency components. Then the signals are digitalized and transferred to baseband with the ADC and proper post-processing in baseband is followed.

Similar to (9.1), we define two signals that represent PA outputs in the two bands of interest

$$\begin{aligned}\tilde{y}(t) &= \tilde{y}_1(t) + \tilde{y}_2(t)\\ \tilde{y}_1(t) &= |y_1(t)|\cos\left(2\pi f_1 t + \angle y_1(t)\right)\\ \tilde{y}_2(t) &= |y_2(t)|\cos\left(2\pi f_2 t + \angle y_2(t)\right),\end{aligned} \tag{9.8}$$

where we assume $f_2 > f_1$. The spectrums for the above signals are illustrated in Fig.9.3, where $B_1$ and $B_2$ denote the respective bandwidths. By feeding the PA output to the mixer shown in Fig.9.3, we can get

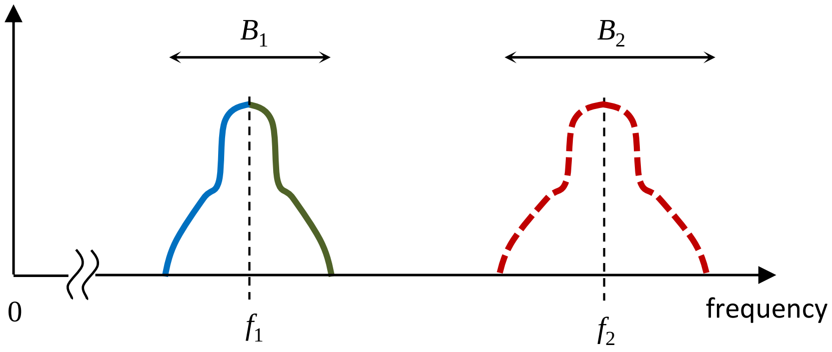


**Fig. 9.3 Illustration of the spectrum for the dual-band PA outputs, where the signals outside the bands of interesting are assumed to be eliminated by matching networks or filters.**

$$\begin{aligned}\tilde{y}_{MIX}(t) &= \tilde{y}_1(t)\cos\left(2\pi f_{lo}t\right)+\tilde{y}_2(t)\cos\left(2\pi f_{lo}t\right)\\ &= \frac{1}{2}\left|y_1(t)\right|\left[\cos\left(2\pi(f_1+f_{lo})t+\angle y_1(t)\right)\right.\\ &\left.+\cos\left(2\pi(f_1-f_{lo})t+\angle y_1(t)\right)\right]+\\ &\frac{1}{2}\left|y_2(t)\right|\left[\cos\left(2\pi(f_2+f_{lo})t+\angle y_2(t)\right)\right.\\ &\left.+\cos\left(2\pi(f_2-f_{lo})t+\angle y_2(t)\right)\right],\end{aligned} \tag{9.9}$$

Where $f_{lo}$ is the frequency of the local oscillator (LO). The LPF filters out undesired signals and only leaves

$$\begin{aligned}\hat{y}_{IF}(t) = &\frac{1}{2}\left|y_1(t)\right|\cos\left(2\pi(f_1-f_{lo})t+\angle y_1(t)\right)+\\ &\frac{1}{2}\left|y_2(t)\right|\cos\left(2\pi(f_2-f_{lo})t+\angle y_2(t)\right).\end{aligned} \tag{9.10}$$

Here, if we set local oscillator frequency in the following range

$$f_{lo} \in \left( f_1, \frac{f_1+f_2}{2} \right], \tag{9.11}$$

we can rewrite (9.10) as

$$\begin{aligned}\hat{y}_{IF}(t) &= \frac{1}{2}\left|y_1(t)\right|\cos\left[-\left(2\pi(f_{lo}-f_1)t-\angle y_1(t)\right)\right]+\\ &\frac{1}{2}\left|y_2(t)\right|\cos\left(2\pi(f_2-f_{lo})t+\angle y_2(t)\right)\\ &=\frac{1}{2}\left|y_1(t)\right|\cos\left(2\pi(f_{lo}-f_1)t-\angle y_1(t)\right)+\\ &\frac{1}{2}\left|y_2(t)\right|\cos\left(2\pi(f_2-f_{lo})t+\angle y_2(t)\right).\end{aligned} \tag{9.12}$$

Note that as $f_{lo}>f_1$, the spectrum in LSB is folded back, and in this case, we can obtain the spectrum plotted in Fig.9.4. Note that as the first term in (9.12) has the inversed phase, the resulted spectrum is a mirrored version of the original signal. Such IF signal can be directly digitalized and separated in the baseband.

By setting $f_{lo}$ to the upper limit in (9.11), we can obtain a spectrum as shown in Fig.9.5, where $y_1(t)$ is folded back and just overlaps with the USB signal. It is still possible to extract the DPD parameters correctly utilizing such a signal, consider that the distortion is properly preserved. The DPD technique introduced in the next section will utilize such signal to reduce the BOM cost paid for the feedback path.

Before dwelling into the exact DPD in detail, it is beneficial to reexamine such signal in Fig.9.5, which can be expressed as

$$\begin{aligned}y_{IF}(t) &= \frac{1}{2}\left|y_1(t)\right|\cos\left(\pi(f_2-f_1)t-\angle y_1(t)\right)\\ &+\frac{1}{2}\left|y_2(t)\right|\cos\left(\pi(f_2-f_1)t+\angle y_2(t)\right),\end{aligned} \tag{9.13}$$

It is also easy to see from the figure that the IF is

$$f_{IF}=\frac{f_2-f_1}{2}, \tag{9.14}$$

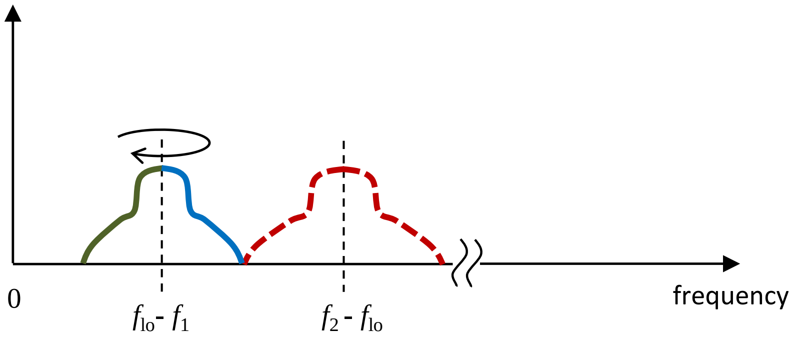


**Fig. 9.4 Illustration of the spectrum of the IF signals, if flo is set in the range defined in (9.11).**

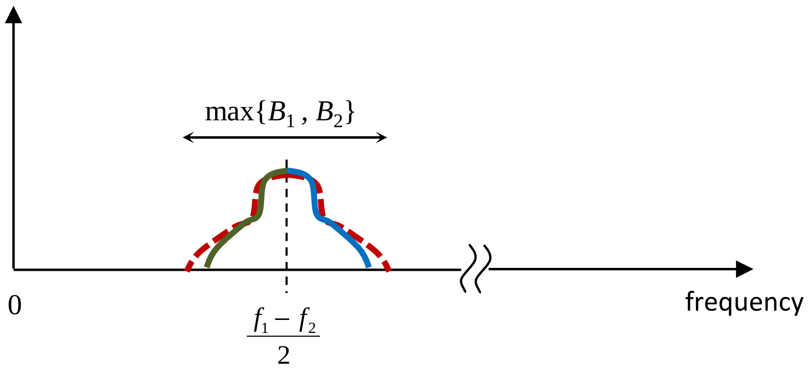


**Fig. 9.5 Illustration of the spectrums of the IF signals for the SFFB architecture, where $f_{lo}$=($f_1$+$f_2$)/2.**

which is one half of the distance of the center frequencies. The above equation implies that if transmitted signals are located in distant bands, it will yield a very high IF, which may be adverse to digitization. To sample the high IF signal, there can be two measures: the first one is to use an extra down converter to convert the signal to lower IF; another way is to utilize bandpass sampling theory to directly digitize the signal. In either case, thereby, the minimum Nyquist rate for the ADC is max(2$B_1$, 2$B_2$). The digitalized signal can be easily transferred to baseband in digital domain with the well-established digital down conversion (DDC) techniques.

The signal in (9.13) can be transferred to baseband and written as:

$$\begin{aligned} y(t) &= \frac{1}{2}\left[|y_1(t)|\exp\left(-j\angle y_1(t)\right) + |y_2(t)|\exp\left(j\angle y_2(t)\right)\right] \\ &= \frac{1}{2}\left[y_1^*(t) + y_2(t)\right] \end{aligned} \tag{9.15}$$

Without losing generality, we can write the feedback signal in baseband as

$$y(t) = y_1^*(t) + y_2(t), \tag{9.16}$$

and which will be utilized to extract the DPDs' coefficients for the respective two bands.

## 9.3.2 Extension to Triple and Multiple Bands

The dual-band SFFB architecture can be easily extended to triple and multiple band signals, and this is accomplished by using the architecture demonstrated in Fig.9.6. More specifically, an $N$-band SFFB architecture contains $N$-1 stages, which are constituted with a series of cascaded mixers and filters. The spectral plots of each stage for a triple-band architecture are illustrated in Fig.9.7 to provide an intuitive vision. The mixer in the first stage employs a LO frequency of ($f_1$+$f_2$)/2, which produces an aliased spectrum at ($f_2$-$f_1$)/2 with the spectrum $S_1$ mirrored. The mixing stage in sequel will further down convert the signals to

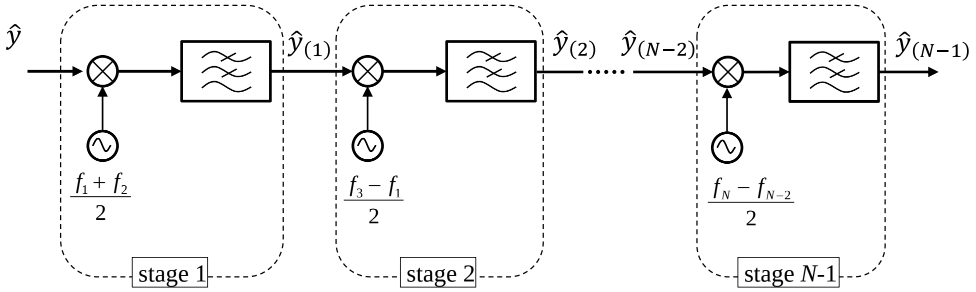


**Fig. 9.6 Illustration of the SFFB architecture for multiple band receiving.**

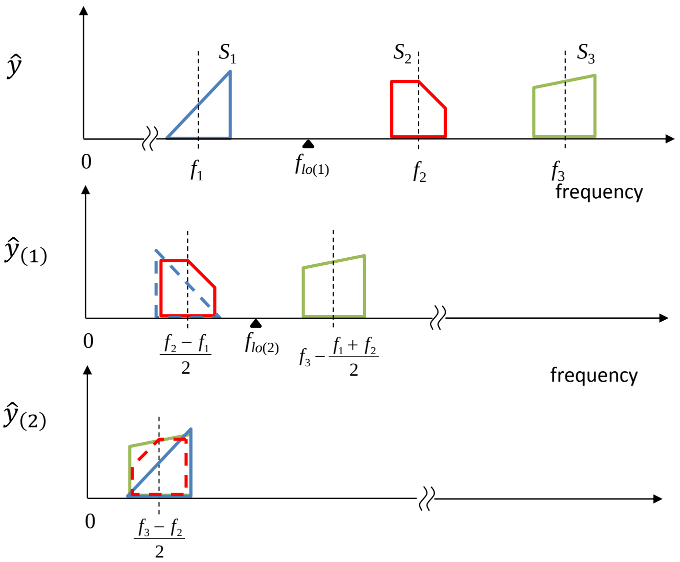


**Fig. 9.7 Illustration of the operation of a triple band SFFB.**

$(f_3+f_2)/2$ and result in the signal $\hat{y}_{(2)}$ as shown in bottom of Fig.9.7, where $S_1$ is reversed back and only $S_2$ is mirrored.

Based on the discussion given above, we can arrive at some conclusions for the $N$-band SFFB architecture. The LO frequency for $i$-th stage can be written as

$$f_{lo(\mathrm{i})}=\begin{cases}\dfrac{f_1+f_2}{2} & i=1\\[2ex] \dfrac{f_{i+1}-f_{i-1}}{2} & i=2,3,\cdots,N-1\end{cases} \tag{9.17}$$

and consequently, the final output becomes

$$\hat{y}_{IF}=(\frac{1}{2})^{N-1}\sum_{i=1}^{N+1}\left|y_i(t)\right|\cos\left[\pi(f_{N+1}-f_N)t+(-1)^{N-i+1}\angle y_i(t)\right]. \tag{9.18}$$

Note that the final IF frequency is only determined by the spacing of $f_N$ and $f_{N+1}$.

## 9.4 DUAL-BAND DIGITAL PREDISTORTION ALGORITHMS FOR SINGLE FEEDBACK PATH ARCHITECTURE

After the introduction of SFFB architecture in last section, the work left is to develop a suitable algorithm to extract the two respective DPDs' coefficients. In this section, an algorithm based on PA modeling and inverse is presented, with its simplicity and suitability to the proposed architecture.

### 9.4.1 SFFB DPD Parameter Estimation

At this point, the available feedback signal is an aliased version of the two signals in respective bands as given in (9.16). Intuitively, separating the two signals becomes a natural consequence before subsequent processing. As we have assumed that the PA inputs and outputs can be related through the bi-variant polynomial given in (9.4), the feedback signal, therefore, can be expressed as the superposition of the outputs of the two polynomials:

$$\begin{aligned}\mathbf{y} &= \mathbf{y}_1^* + \mathbf{y}_2 \\ &= \left(K(\mathbf{v}_1, \mathbf{v}_2)\mathbf{A}\right)^* + K(\mathbf{v}_2, \mathbf{v}_1)\mathbf{B} \\ &= \left(\mathbf{K}_1\mathbf{A}\right)^* + \mathbf{K}_2\mathbf{B},\end{aligned} \tag{9.19}$$

where $\mathbf{A},\mathbf{B} \in \mathbb{C}^{\frac{MQ(Q+1)}{2}}$ are the respective coefficient vectors to be determined, and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{y}_1, \mathbf{y}_2 \in \mathbb{C}^L$ are signal vectors. Here, in the sake of compact denote, we use $\mathbf{K}_1$ and $\mathbf{K}_2$ to replace $K(\mathbf{v}_1, \mathbf{v}_2)$ and $K(\mathbf{v}_2, \mathbf{v}_1)$. As the coefficients are linear respect to the kernels, the above equation can be written as

$$\mathbf{y} = \begin{bmatrix}\mathbf{K}_1^* & \mathbf{K}_2\end{bmatrix}\begin{bmatrix}\mathbf{A}^* \\ \mathbf{B}\end{bmatrix}. \tag{9.20}$$

The well-established least squares (LS) can be employed to estimate the coefficients:

$$\begin{aligned}\begin{bmatrix}\mathbf{A}^* \\ \mathbf{B}\end{bmatrix} &= \left(\begin{bmatrix}\mathbf{K}_1^* & \mathbf{K}_2\end{bmatrix}^H\begin{bmatrix}\mathbf{K}_1^* & \mathbf{K}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{K}_1^* & \mathbf{K}_2\end{bmatrix}^H\mathbf{y} \\ &= \left(\begin{bmatrix}\mathbf{K}_1^T \\ \mathbf{K}_2^H\end{bmatrix}\begin{bmatrix}\mathbf{K}_1^* & \mathbf{K}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{K}_1^T \\ \mathbf{K}_2^H\end{bmatrix}\mathbf{y} \\ &= \left(\begin{bmatrix}\mathbf{K}_1^T\mathbf{K}_1^* & \mathbf{K}_1^T\mathbf{K}_2 \\ \mathbf{K}_2^H\mathbf{K}_1^* & \mathbf{K}_2^H\mathbf{K}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{K}_1^T \\ \mathbf{K}_2^H\end{bmatrix}\mathbf{y}.\end{aligned} \tag{9.21}$$

with **A** and **B** in hand, the respective PA output signals can be produced with the models:

$$\begin{aligned}\mathbf{y}_{m1} &= K(\mathbf{x}_1, \mathbf{x}_2)\mathbf{A} \\ \mathbf{y}_{m2} &= K(\mathbf{x}_2, \mathbf{x}_1)\mathbf{B}.\end{aligned} \tag{9.22}$$

Then the work left is to utilize (9.7) to estimate the DPD coefficients:

$$\begin{aligned}\mathbf{W}_1 &= \left[K^H(\mathbf{y}_{m1}, \mathbf{y}_{m2})K\ (\mathbf{y}_{m1}, \mathbf{y}_{m2})\right]^{-1} K^H(\mathbf{y}_{m1}, \mathbf{y}_{m2})\mathbf{x}_1 \\ \mathbf{W}_2 &= \left[K^H(\mathbf{y}_{m2}, \mathbf{y}_{m1})K\ (\mathbf{y}_{m2}, \mathbf{y}_{m1})\right]^{-1} K^H(\mathbf{y}_{m2}, \mathbf{y}_{m1})\mathbf{x}_2.\end{aligned} \tag{9.23}$$

It is easy to see that the above procedure contains two steps: 1) PA model coefficients extraction using equation (9.21), and produce the PA output with achieved coefficients following (9.22). 2) DPD coefficients estimation, which can be seen as a procedure of inverse modeling.

## 9.4.2 Discussion on the Proposed Method

After establishing of the algorithm, it is interesting to examine the computational complexity it brings about. In the conventional method, two times of $L$-sample $N_k$-dimensional LS ($N_k$ is the number of model coefficients) are employed to extract the DPD coefficients, while the proposed method needs an additional LS process to estimate the model coefficients. Standard LS involve three times of matrix multiplication and one time of matrix inverse. Generally, QR (or Cholesky) factorization is used to achieve numerically stable inverse, so that the conventional method's complexity is twice of $O(N_kL)+2O(N_k^2L)+O(N_k^3)$. The proposed method, however, with an additional $L$-sample $2N_k$-dimensional LS process, results in additional complexity of $2O(N_kL)+8O(N_k^2L)+8O(N_k^3)$.

The increase in computational complexity would be rather worthful, considering the significant reduction in BOM cost. It is worth to mention that, SFFB DPD also allows the use of sample-by-sample adaptive algorithms, such as least mean square (LMS) and recursive least square (RLS), which have lower computational complexities, to adopt it to some special applications that need rapid adaption in order to correspond to fast signal status variation [22]. Such algorithms can be easily derived and are not discussed here, due to the limited length.

## 9.4.3 Practical Implementation Issues

Implementing with low-distortion peripheral circuits (frequency converters, data converters, etc.) is the key-to-success for maximizing the DPD performance. In most practical applications, this is difficult to ensure and it is always necessary to recourse to digital signal processing techniques to facilitate post-compensations for these circuits.

An important issue is that the peripheral analog circuits, such as mixers and filters, generally show unequal magnitude-frequency and phase-frequency responses respect to the two signals in different bands, see Fig.9.8. For dual-band implementations with small carrier spacing, the two signals have similar gain and phase shift as the response in a small fre

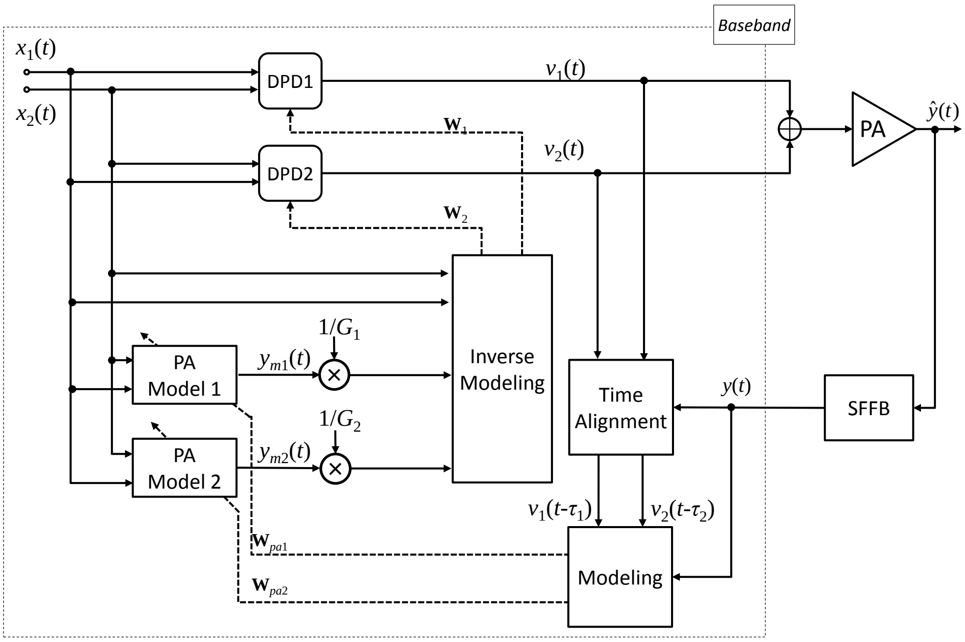


**Fig. 9.8 Detailed diagram of the proposed SFFB architecture for concurrent dual-band predistorter.**

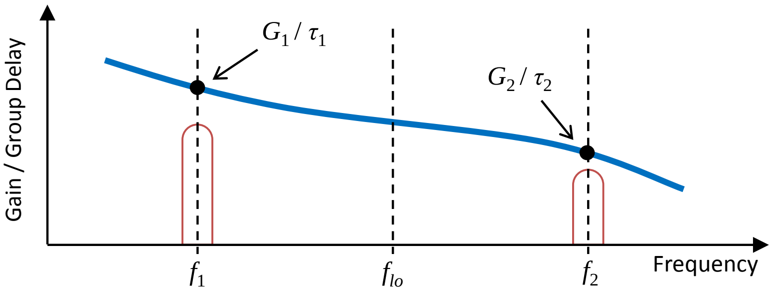


**Fig. 9.9 Illustration of the effect of nonideal frequency response.**

quency range is similar. Conversely, for the signals located in distant bands, the circuits can cause misalignment in power and phase, which hinder the estimation of optimized DPD parameters.

The unequal magnitude-frequency can be interpreted as different gains for respective bands and it can be easily calibrated with a continuous wave (CW) test. The uneven phase-frequency response, on the other hand, bestows dissimilar phase shifts on the two signals, which directly transfers to non-constant group-delays, so that the two signals are mismatched in time.

It should be noted that the feedback signal actually contains the contribution of the circuits in both transmission and feedback paths, while the post-compensation of unequal gain should only be taken place for the feedback path circuits. This is to maintain the gain variation for the transmission path as the DPD should never alter the PA output power. Also, the nonidealities caused by transmission path can be interpreted as part of the distortion caused by PA and is to be compensated by the DPD.

To show the aforementioned issue we assume the actual feedback signal is given by

$$y(t)=G_1 y_1^*(t-\tau_1)+G_1 y_2(t-\tau_2), \tag{9.24}$$

where $\tau_1$ and $\tau_2$ denote the delays respect to the input signal $x_1(t)$ and $x_2(t)$. A commonly used method for delay estimation is to find the delay value that maximizes or minimizes an extra correlation function. For instance the cross correlation function can be used here, and we have

$$\begin{aligned}\tau_1 &= \max_{\tau}\left|xcorr\left(x_1(t-\tau), y^*(t)\right)\right| \\ \tau_2 &= \max_{\tau}\left|xcorr\left(x_2(t-\tau), y(t)\right)\right|,\end{aligned} \tag{9.25}$$

where *xcorr* denotes the cross-correlation of two sequences. Note that actual $\tau_1$ and $\tau_2$ may be fractional number that indicating that the delays smaller than a sampling period. Some extra methods, such as the one introduced in [99], can be used to determine the fractional delay, which is not further discussed here.

With the estimated delays in hand, we can use $v_1(t\text{-}\tau_1)$ and $v_2(t\text{-}\tau_2)$ to replace the corresponding signals in (9.19) and (9.21) to calculate the model parameters. Finally, the respective PA output signals are given as

$$\begin{aligned}\mathbf{y}_{m1} &= \frac{K(\mathbf{x}_1,\mathbf{x}_2)\mathbf{A}}{G_1} \\ \mathbf{y}_{m2} &= \frac{K(\mathbf{x}_2,\mathbf{x}_1)\mathbf{B}}{G_2}.\end{aligned} \tag{9.26}$$

The work left is the same as previous introduced.

## 9.5 EXPERIMENTAL RESULTS

### 9.5.1 Description of the Test Bed

Experiments were taken place to evaluate the suitability of the proposed DPD. The measurement applied a general purposed wideband PA HMC-C074 from Hittite corporation, which is driven by R&K AA160. The nonlinear nature of the amplifier is apparent in the results of CW tests shown in Fig.9.10, from which it can be seen that the PA shows apparently dissimilar characteristics at the frequencies of 1.75 GHz and 2.75 GHz. More specifically, the PA has around 1.8 dB variation for linear gain and 1 dB variation for output saturation power for the two frequencies. As already mentioned before, the DPD should maintain the linear gain for the respective frequencies, more specifically, the desired gains for LSB and USB should keep a 1.8 dB variation.

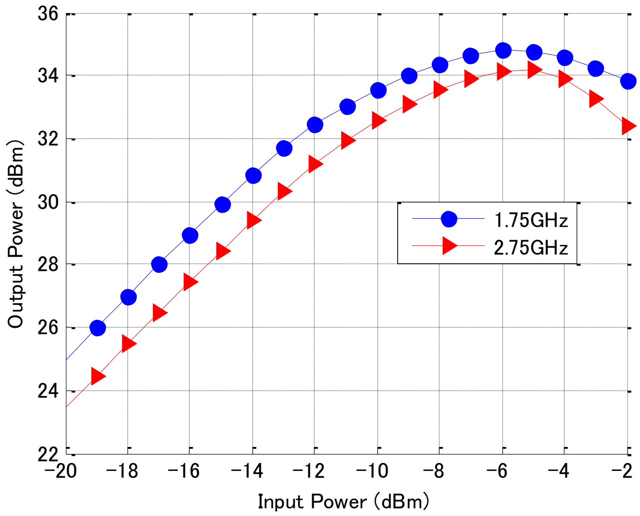


**Fig. 9.10 Measured swept power data of the amplifier in test which were obtained in 1.75 GHz and 2.75 GHz.**

The test bed (see Fig.9.11) is comprised of an arbitrary signal generator (AWG) M8190A and a PXIe signal analyzer, both from Agilent, as the transmitter and receiver, respectively. The M8190A, which has a sampling rate of maximum 12 GSa/s and an analog bandwidth of 5 GHz, can directly output the dual-band RF signals. The PXIe contains a broadband mixer M9362A-D01 to down convert the frequency to the desired IF. A signal generator N5183 is used as the local oscillator (LO). The mixer is followed by a variable gain IF amplifier M9352A, which is used to adjust the IF signal's peak-to-peak voltage to fully use every bit of the downstream digitizer M9202A as much as possible. A clock synthesizer M9300A provides the clock for the digitizer and other reference signals needed for phase locking. The digitalized signal is processed by the Agilent VSA81900 software, which serves as a digital down convertor (DDC) that transfers IF signal to baseband.

The magnitude-frequency response for PXIe is plotted in Fig.9.12, where the LO frequency $f_{lo}$ was fixed to (1.75+2.75)/2=2.25 GHz and the input frequency $f_c$ was swept. The frequency at $f_c$=$f_{lo}$ is heavily attenuated as the mixer output is AC-coupled. It can be inspected from the figure that, for instance, albeit well designed, M9362A-D01 still shows a gain variation over 2 dB for the frequency spacing of 1 GHz.

## 9.5.2 Test with LTE-Advanced Signals

The first measurement scenario applies LTE signals with 10 MHz (9.015MHz for occupied bandwidth) and 20 MHz (18.015MHz for occupied bandwidth) bandwidth for the LSB and USB, which are located at 1.75 GHz and 2.75 MHz. Each LTE signal has a peak-to-average power ratio (PAPR) of around 10 dB, which was reduced to 7 dB with crest factor reduction (CFR) in order to maximize the DPD performance. Experiment with the conventional two-branch architecture was also carried out to give a comparison to the proposed single feedback path technique. In the test, another vector signal analyzer Rohde & Schwarz

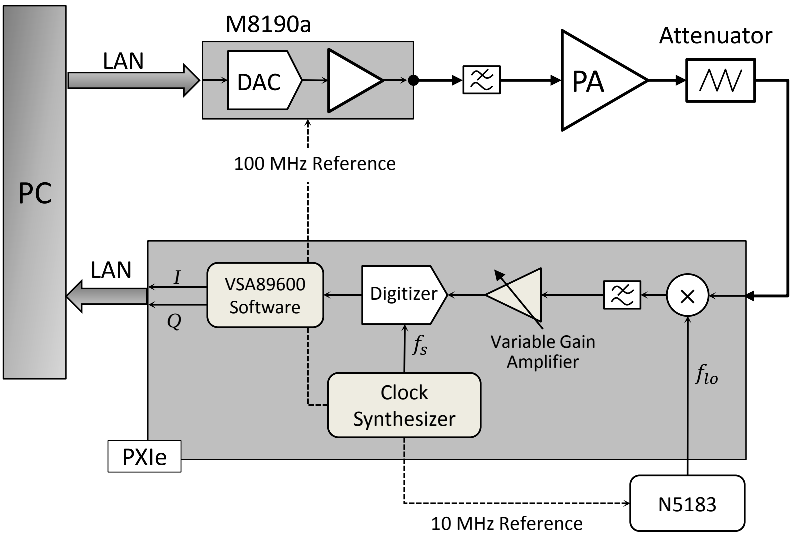


**Fig. 9.11 Diagram of the test bed.**

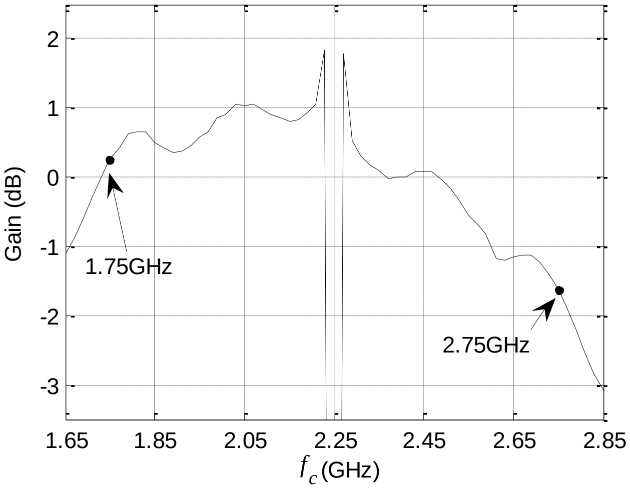
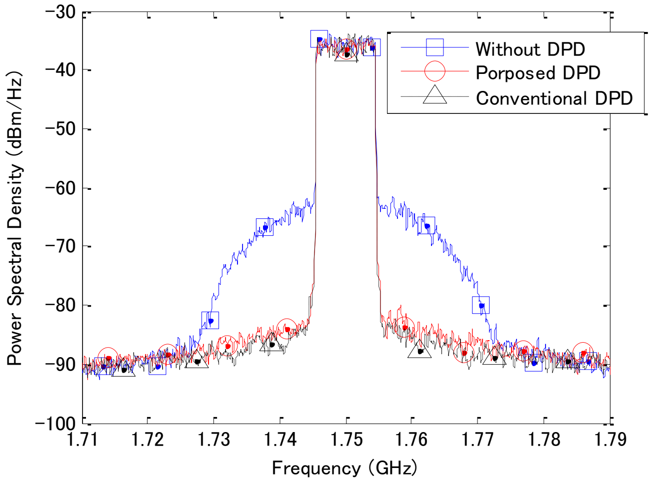


**Fig. 9.12 Measured magnitude-frequency response of the mixer, where the LO frequency was set to 2.25 GHz.**

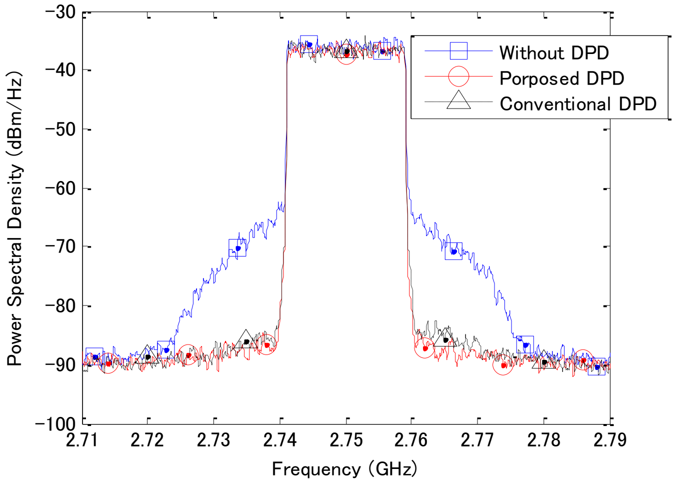
FSW, which has an IF bandwidth of 320 MHz, was used accompanying with the PXIe, in order to receive the signals in the two bands simultaneously.

It is clear from Fig.9.13 that the single feedback path architecture and conventional dual-branch architecture have almost the same out-of-band compensation performance. This has proved the advancement of the proposed architecture: lower hardware overhead and non-sacrificed performance.

The measured error vector magnitude (EVM), adjacent channel leakage ratio (ACLR) and normalized mean square error (NMSE), with and without proposed DPD, are summarized in Table 9-1 and Table 9-2. Note that the EVM is a figure of merit benchmarking the error of the whole system, and the test bed (instruments) and CFR also introduces error to a



(a) LSB output spectrum for scenario 1, 10 MHz LTE signal.



(a) USB output spectrum for scenario 1, 20 MHz LTE signal.

**Fig. 9.13 Measured spectral plots: without DPD, with conventional dual-branch DPD and proposed single feedback path DPD.**

certain extent. It can be seen by inspecting Table I that the instruments gives an EVM of 0.04% (-64.4 dB) for 10MHz signal and which is degraded by about 0.05% for the 20MHz signal. The CFR, however, introduces apparent distortions, and the EVMs for respective signals are 3.47% (-29.2 dB) and 3.36% (-29.4 dB). The PA, furthermore, adds another 3~4% degradation in EVM and its effect can be perfectly compensated by the DPD.

It can be seen from Table I that the CFR has set the lower bound of the EVM that the DPD can achieve. In order to barely evaluate the effect of DPD, the adjacent channel leakage ratio (ACLR) and normalized mean square error (NMSE) are measured and summarized in Table 9-2. The ACLRs were improved by 22 dB and 19 dB for the respective bands with the effect of DPD. On the other hand, the DPD also gives evident improvements in NMSE,

Table 9-1 Measured EVMs

(For LTE-Advanced signals with multiple CCs, the EVM is averaged)

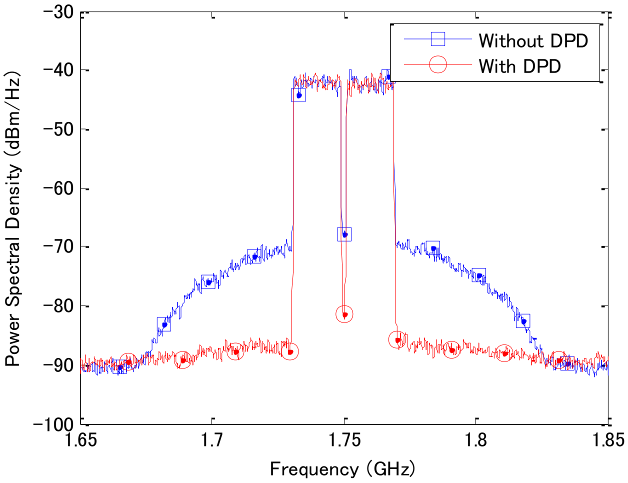
| Scenario | Test Signals | Test Bed | CFR | CFR +PA | CFR +DPD +PA |
|---|---|---|---|---|---|
| 1 | 10MHz LTE | 0.06% | 3.47% | 6.78% | 3.51% |
|  | 20MHz LTE | 0.11% | 3.36% | 7.25% | 3.41% |
| 2 | 20MHz LTE×2 | 0.18% | 3.84% | 7.56% | 3.93% |
|  | 20MHz LTE×3 | 0.21% | 3.63% | 8.18% | 3.74% |
| 3 | 20MHz LTE×1 | 0.11% | 3.36% | 7.38% | 3.56% |
|  | 20MHz LTE×4 | 0.23% | 4.07% | 8.48% | 4.17% |

Table 9-2 Measured NMSEs and ACLRs

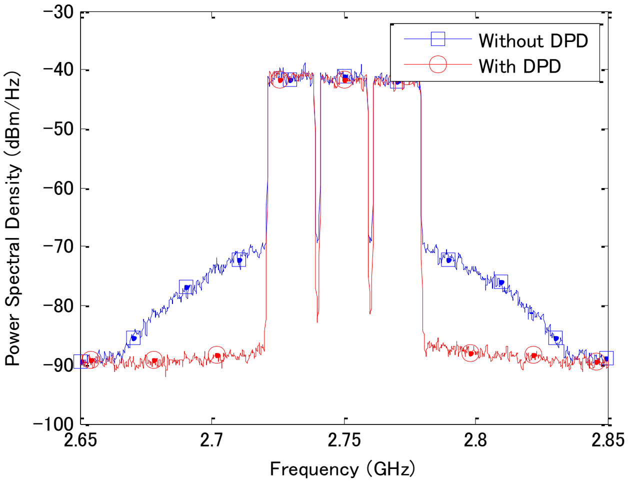
| Scenario | Test Signals |  | NMSE (dB) | ACLR (dB) |  |
|---|---|---|---|---|---|
|  |  |  |  | lower | upper |
| 1 | 10MHz LTE | Without DPD | -20.28 | -28.58 | -28.85 |
|  |  | With DPD | -39.17 | -50.53 | -50.16 |
|  | 20MHz LTE | Without DPD | -22.97 | -33.09 | -33.40 |
|  |  | With DPD | -41.88 | -52.53 | -52.72 |
| 2 | 20MHz LTE×2 | Without DPD | -20.83 | -27.90 | -28.02 |
|  |  | With DPD | -37.34 | -47.75 | -47.28 |
|  | 20MHz LTE×3 | Without DPD | -21.98 | -29.73 | -30.89 |
|  |  | With DPD | -38.78 | -48.79 | -48.78 |
| 3 | 20MHz LTE | Without DPD | -19.14 | -28.71 | -29.21 |
|  |  | With DPD | -37.82 | -47.89 | -47.25 |
|  | 20MHz LTE×4 | Without DPD | -21.46 | -29.23 | -31.28 |
|  |  | With DPD | -35.45 | -45.55 | -46.28 |

which is calculated by comparing the signal after CFR and PA output.

In scenario 2, two 20 MHz carrier components (CCs) were aggregated at the 1.75 GHz band and another three contiguous CCs were located at 2.75 GHz. The resulting spectral plots for the both bands that depicted in Fig.9.14 confirm the validity of the proposed DPD. It can be seen from the figure that the out-of-band emission can be significantly suppressed. Applying the SFFB DPD enables the improvement of ACPRs by more than 20 dB in each



(a) LSB output spectrum for scenario II, 2-CC LTE signal.

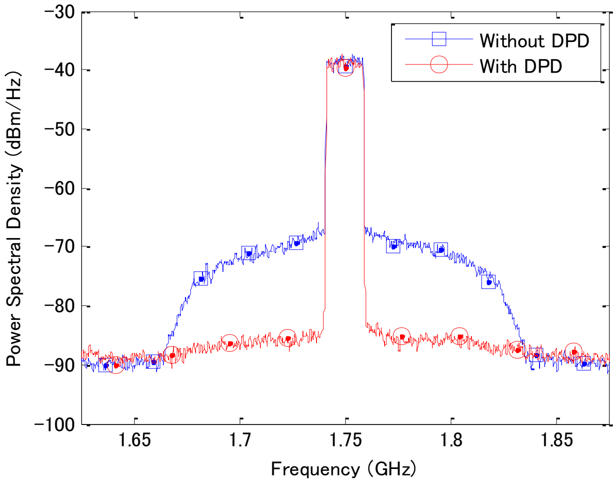


(b) USB output spectrum for scenario II, 3-CC LTE signal.

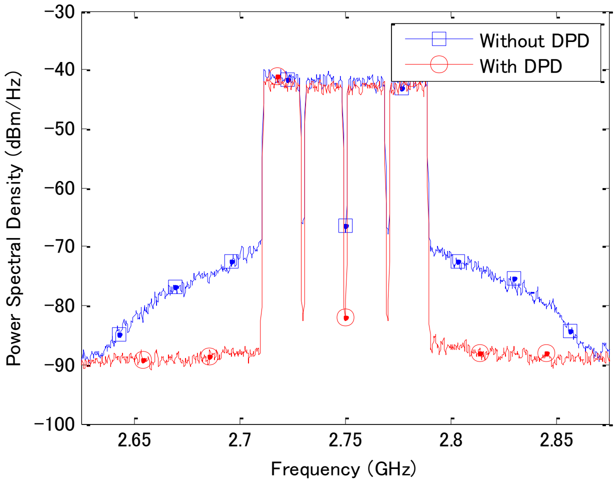
**Fig. 9.14 Measured spectral plots with and without DPD for Scheme 2**

band and ACPRs are lower than -47 dB in both bands, as shown in Table II.

Scenario 3 assumes a 1-CC LTE for the LSB and a 4-CC LTE for the USB. Thus, the signal in the USB has a contiguous bandwidth of 80 MHz. Note that the signal in LSB has a very wideband out-of-band distortion, which is mainly generated from the cross-band modulations. The USB signal shows an obvious unbalance in the lower and upper band ACLR owing to the strong memory effect it suffers. From Fig.9.15 we can inspect the significant improvement of out-of-band emission for both LSB and USB signals. The improvement in EVMs is around 4% for both bands and the improvements in NMSEs are 18 dB and 14 dB for the respective bands.



(a) LSB output spectrum for scenario II, 1-CC LTE signal.



(b) USB output spectrum for scenario II, 4-CC LTE signal.

**Fig. 9.15 Measured spectral plots with and without DPD for Scheme 3**

# Chapter 10.

# Conclusions and Future Works

## 10.1 CONCLUSIONS

The work presented in this thesis was motivated by the fact that the performance of the wireless communication systems is affected by the nonlinearities of the actual circuits to a large extent. The power amplifier and low noise amplifier are considered as they are the primary source of nonlinearity. The counter strategies taken in this thesis are DPD and PNLC, as they are a good balance for cost and performance.

Chapter 1 provided the principal motivation for the work, which is firmly connected with the current trend of wireless communication system development. A thesis outline and a list of contributions was given.

Chapter 2 has given a brief introduction of the basics about digital signal and nonlinear amplifiers. The distortion mechanism of the amplifier, in terms of AM-AM/ AM-PM conversions and memory effect are presented.

Chapter 3 sets up the mathematical formulations for the nonlinear models, including mathematical expression and linear architecture of the polynomial model. It is followed by the presen-tation of a general mathematical method for inverting the linear model as well as the method for obtaining the inverse of a given polynomial.

Chapter 4 describes practical issues regarding instrumental experiment and FPGA implementation. Some issues that, such as I/Q impairment and gain/delay alignment, must to be resolved in practice are discussed.

In chapter 5, an adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters was presented. It was successfully implemented in an FPGA without using any other auxiliary processor. A simplified multiplier-free NLMS algorithm, especially suitable for FPGA implementation, for fast adapting the LUT was proposed. Many design methodologies and practical implementation issues were discussed in details. Experimental results have shown that the DPD performed robustly when it is involved in the multi-channel transmitter.

In chapter 6, a digital predistortion technique utilizing spectral extrapolation has been proposed. With band-limited feedback signal, the requirement on ADC speed can be significantly released. Experimental results have validated the feasibility of the proposed technique for coping with band-limited feedback signal. It has been shown that adequate linearization performance can be achieved even if the acquisition bandwidth is less than the original signal bandwidth. The experimental results obtained by using LTE-Advanced signal of 320 MHz bandwidth are quite satisfactory, and to the authors' knowledge, this is the first high-performance wideband DPD ever been reported.

In chapter 7, we proposed and demonstrated a novel blind nonlinear compensation technique to mitigate the effects of intermodulation distortion in receivers. The compensator was implemented with an FPGA, and experimental results have justified its suitability to combat the nonlinearity of the real LNA.

In this chapter 8, an adaptive technique for successively compensating the nonlinear distortion of the tunable concurrent dual-band receiver was proposed. The mechanism of the nonlinear behavior of the receiver was analyzed by assuming two band-limiting signals in different frequency bands, and a corresponding compensator utilizing linear term de-embedding was introduced. In the experimental measurement with an actual amplifier, the proposed compensator yielded significant improvement in EVMs. It is also worth to mention that the blind compensation may not be applicable to nonlinear models with memory. This is probably because the cost function is based on the out-of-band emission, which lacks of the information of the frequency band flatness. However, this can be a future work for blind nonlinear compensations.

In chapter 9, a spectra-folding feedback architecture for linearization of concurrent dual-band PAs is proposed. Compared with conventional dual-band architecture, the BOM cost of the feedback path is halved. In this architecture, simple modifications are made to the down converter while the operation of the ADC is kept unchanged. With the proposed algorithm for DPD parameter estimation in the baseband, the SFFB DPD can achieve comparative performance as the conventional dual-branch DPD, as proved by experiment with LTE/LTE-Advanced signals. Although the proposed architecture may not be seen as very groundbreaking, it is rather promising and we strongly believe that it can replace the existing dual-branch architecture of dual-band predistortion in the near future.

To conclude, the works done in this thesis are based on digital processing of the baseband signal, with a primary focus on the practical implementations and complexity, for nonlinear compensation of both transmitters and receivers. To the best of our knowledge, there exists no other similar work in academic word. Thus, in this sense, the research we have done here can be seen as rather ground-breaking. In a wider context, we also believe that bringing the digital signal processing closer to the radio systems to reduce the difficulties of implementations and cut down the cost of circuits would be very meaningful. Therefore, in addition to some FPGA implementation and instrumental measurements, the techniques proposed in this thesis are rather practically oriented.

## 10.2 FUTURE WORKS

Many works have been done focusing on nonlinear compensation for transmitters and abundant literatures come out especially for DPD techniques. However, the real challenges for wireless communications are raised by designing the receiver front-ends, as which are more delicate to interference, noise and hardware impairments. To the best of the author's knowledge, there are only a few of researches have been taken place to address the nonlinear problem for receivers compare to the large amount of papers discussing DPD. Furthermore, the signals arrives at the receiver have time varying average power in realistic environment, owing to the fading effect. Therefore, the receiver front-end linearization should handle the dynamic power changing and track the signal's status variation.

On the other hand, the research on DPD has turned to the practical implementations with minimized hardware cost. This is motivated by the adoption of heterogeneous networks therein many small cells are deployed on top of macro cells.

In addition, to enhance the coverage and reduce the cost, heterogeneous networks uses less baseband equipment (BBE), which are complex and expensive, and implements front-ends in the remote antenna units (RAUs) or remote radio units (RRUs). One BBE is connected with multiple RAUs and RRUs through optical fibers, thereby electrical/optical (E/O) and optical/electrical (O/E) converters are employed in downlink and uplink signal transmission. The nonlinearity of the E/O converter is inevitable for higher frequencies, therein DPD comes out to be the best solution for downlink linearization in such systems. In such sense, the usage of DPD can be extended rather than single-purpose PA linearization

# References

[1] Cripps Steve, Advanced Techniques in RF Power Amplifier Design (Artech House, 2002)

[2] Cripps Steve, RF Power Amplifiers for Wireless Communications (Artech House, 2006)

[3] Terzian A Ps, D B Clark, and R W Waugh, “Broad-Band GaAs Monolithic Amplifier Using Negative Feedback,” *Microwave Theory and Techniques, IEEE Transactions on*, 30 (1982), 2017–2020

[4] Hyde J, “Restrictions on Switching in Positive-Negative Feedback Control Systems,” Automatic Control, IEEE Transactions on, 15 (1970), 502–503

[5] Ezzeddine A K, H.-L.A. Hung, and H.-C. Huang, “An MMAC C-Band FET Feedback Power Amplifier,” *Microwave Theory and Techniques, IEEE Transactions on*, 38 (1990), 350–357

[6] Coskun A.Hakan, and S Demir, “A Mathematical Characterization and Analysis of a Feedforward Circuit for CDMA Applications,” *Microwave Theory and Techniques, IEEE Transactions on*, 51 (2003), 767–777

[7] Coskun A Hakan, A Mutlu, and S Demir, “A Multitone Model of Complex Enveloped Signals and Its Application in Feedforward Circuit Analysis,” *Microwave Theory and Techniques, IEEE Transactions on*, 53 (2005), 2171–2178

[8] Danyuk D L, and G V Pilko, “Feedforward Amplifiers Incorporate Parallel Output Summing,” *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 41 (1994), 912–915

[9] Hassin D, and R Vahldieck, “Feedforward Linearization of Analog Modulated Laser Diodes-Theoretical Analysis and Experimental Verification,” *Microwave Theory and Techniques, IEEE Transactions on*, 41 (1993), 2376–2382

[10] Kahn L R, “Comparison of Linear Single-Sideband Transmitters with Envelope Elimination and Restoration Single-Sideband Transmitters,” Proceedings of the IRE, 44 (1956), 1706–1712

[11] Kahn L R, “Single-Sideband Transmission by Envelope Elimination and Restoration,” Proceedings of the IRE, 40 (1952), 803–806

[12] Raab F H, “Intermodulation Distortion in Kahn-Technique Transmitters,” *Microwave Theory and Techniques, IEEE Transactions on*, 44 (1996), 2273–2278

[13] Cox D C, and R Leck, “A VHF Implementation of a LINC Amplifier,” *Communications, IEEE Transactions on*, 24 (1976), 1018–1022

[14] Hetzel S A, A Bateman, and J P McGeehan, “LINC Transmitter,” Electronics Letters, 27 (1991), 844–846

[15] Rustako A J, and Y S Yeh, “A Wide-Band Phase-Feedback Inverse-Sine Phase Modulator with Application Toward a LINC Amplifier,” *Communications, IEEE Transactions on*, 24 (1976), 1139–1143

[16] Temerinac M, and B Edler, “LINC: a Common Theory of Transform and Subband Coding,” *Communications, IEEE Transactions on*, 41 (1993), 266–274

[17] Nagata Yoshinori, “Linear Amplification Technique for Digital Mobile Communications,” in Vehicular Technology Conference, 1989, IEEE 39th, 1989, pp. 159–164

[18] Cavers James K, “Amplifier Linearization Using a Digital Predistorter with Fast Adaptation and Low Memory Requirements,” *Vehicular Technology, IEEE Transactions on*, 39 (1990), 374–382

[19] Jeon Won Gi, Kyung Hi Chang, and Yong Soo Cho, “An Adaptive Data Predistorter for Compensation of Nonlinear Distortion in OFDM Systems,” *IEEE Transactions on Communications*, 45 (1997), 1167–1171

[20] Ma Yuelin, Songbai He, Yoshihiko Akaiwa, and Yasushi Yamao, “An Open-Loop Digital Predistorter Based on Memory Polynomial Inverses for Linearization of RF Power Amplifier,” *International Journal of RF and Microwave Computer-Aided Engineering*, 21 (2011), 589–595

[21] Ma Yuelin, Yasushi Yamao, Yoshihiko Akaiwa, and Koji Ishibashi, “Wideband Digital Predistortion Using Spectral Extrapolation of Band-Limited Feedback Signal,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61 (2014), 2088–2097

[22] Ma Yuelin, Yasushi Yamao, Yoshihiko Akaiwa, and Chunlei Yu, “FPGA Implementation of Adaptive Digital Predistorter With Fast Convergence Rate and Low Complexity for Multi-Channel Transmitters,” *IEEE Transactions on Microwave Theory and Techniques*, 61 (2013), 3961–3973

[23] Akasaki Takeshi, Motonori Iwata, and Yoshihiko Akaiwa, “A Mathematical Expression of Nonlinear Distortion in RF Power Amplifier,” in IEEE Vehicular Technology Conference, 2004, pp. 4217–4220

[24] Kim J, and K Konstantinou, “Digital Predistortion of Wideband Signals Based on Power Amplifier Model with Memory,” Electronics Letters, 37 (2001), 1417–1418

[25] Morgan Dennis R, Zhengxiang Ma, Jaehyeong Kim, Michael G Zierdt, and John Pastalan, “A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers,” *Signal Processing, IEEE Transactions on*, 54 (2006), 3852–3860

[26] Raich Raviv, and G Tong Zhou, “On the Modeling of Memory Nonlinear Effects of Power Amplifiers for Communication Applications,” in Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. Proceedings of 2002 IEEE 10th, 2002, pp. 7–10

[27] Zhu Anding, and Thomas J Brazil, “Behavioral Modeling of RF Power Amplifiers Based on Pruned Volterra Series,” Microwave and Wireless Components Letters, IEEE, 14 (2004), 563–565

[28] Watkins Bruce E, and Richard North, “Predistortion of Nonlinear Amplifiers Using Neural Networks,” in Military Communications Conference, 1996. MILCOM’96, Conference Proceedings, IEEE, 1996, i, 316–320

[29] Zhang Qi-Jun, Kuldip C Gupta, and Vijay K Devabhaktuni, “Artificial Neural Networks for RF and Microwave Design-from Theory to Practice,” *Microwave Theory and Techniques, IEEE Transactions on*, 51 (2003), 1339–1350

[30] Ma Yuelin, and Yasushi Yamao, “Blind Nonlinear Compensation Technique for RF Receiver Front-End,” in 43rd European Microwave Conference (Nuremberg: IEEE, 2013), pp. 1527–1530

[31] Ma Yuelin, Yasushi Yamao, Koji Ishibashi, and Yoshihiko Akaiwa, “Adaptive Compensation of Inter-Band Modulation Distortion for Tunable Concurrent Dual-Band Receivers,” *IEEE Transactions on Microwave Theory and Techniques*, 61 (2013), 4209–4219

[32] Razavi Behzad, “Design Considerations for Direct-Conversion Receivers,” Circuits and Systems II: *Analog and Digital Signal Processing, IEEE Transactions on*, 44 (1997), 428–435

[33] Rudolph Matthias, Peter Heymann, and Hermann Boss, “Impact of Receiver Bandwidth and Nonlinearity on Noise Measurement Methods [Application Note],” *Microwave Magazine, IEEE*, 11 (2010), 110–121

[34] Bevilacqua Andrea, and Francesco Svelto, “Statistical Analysis of Second-Order Intermodulation Distortion in WCDMA Direct Conversion Receivers,” *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 52 (2005), 117–121

[35] Kahrizi Masoud, Jaleh Komaili, John E Vasa, and Darioush Agahi, “Adaptive Filtering Using LMS for Digital TX IM2 Cancellation in WCDMA Receiver,” in Radio and Wireless Symposium, 2008 IEEE, 2008, pp. 519–522

[36] Keehr Edward A, and Ali Hajimiri, “Equalization of Third-Order Intermodulation Products in Wideband Direct Conversion Receivers,” *Solid-State Circuits, IEEE Journal of*, 43 (2008), 2853–2867

[37] Keehr Edward A, and Ali Hajimiri, “Successive Regeneration and Adaptive Cancellation of Higher Order Intermodulation Products in RF Receivers,” *Microwave Theory and Techniques,*

*IEEE Transactions on*, 59 (2011), 1379–1396

[38] Lou Shuzuo, and Howard C Luong, “A Linearization Technique for RF Receiver Front-End Using Second-Order-Intermodulation Injection,” *Solid-State Circuits, IEEE Journal of,* 43 (2008), 2404–2412

[39] Habibi Hooman, Erwin J G Janssen, W Yan, and Jan W M Bergmans, “Digital Compensation of Cross-Modulation Distortion in Multimode Transceivers,” *IET communications*, 6 (2012), 1724–1733

[40] Hueber Gernot, Yaning Zou, Krzysztof Dufrene, Rainer Stuhlberger, and Mikko Valkama, “Smart Front-End Signal Processing for Advanced Wireless Receivers,” *Selected Topics in Signal Processing, IEEE Journal of,* 3 (2009), 472–487

[41] Valkama Mikko, A Shahed Hagh Ghadam, Lauri Anttila, and Markku Renfors, “Advanced Digital Signal Processing Techniques for Compensation of Nonlinear Distortion in Wideband Multicarrier Radio Receivers,” *Microwave Theory and Techniques, IEEE Transactions on*, 54 (2006), 2356–2366

[42] Zou Qiyue, Mohyee Mikhemar, and Ali H Sayed, “Digital Compensation of Cross-Modulation Distortion in Software-Defined Radios,” *Selected Topics in Signal Processing, IEEE Journal of,* 3 (2009), 348–361

[43] Dahlman Erik, Stefan Parkvall, and Johan Skold, 4G: LTE/LTE-Advanced for Mobile Broadband (Academic Press, 2013)

[44] Andrews J G, S Buzzi, Wan Choi, S V Hanly, A Lozano, A C K Soong, and others, “What Will 5G Be?,” *Selected Areas in Communications, IEEE Journal on*, 32 (2014), 1065–1082

[45] Gohil Asvin, Hardik Modi, and Shobhit K Patel, “5G Technology of Mobile Communication: A Survey,” in Intelligent Systems and Signal Processing (ISSP), 2013 International Conference on, 2013, pp. 288–292

[46] Wang Y, J Li, L Huang, Y Jing, A Georgakopoulos, and P Demestichas, “5G Mobile: Spectrum Broadening to Higher-Frequency Bands to Support High Data Rates,” *Vehicular Technology Magazine, IEEE*, 9 (2014), 39–46

[47] Ma Yuelin, Yasushi Yamao, Yoshihiko Akaiwa, and Koji Ishibashi, “Wideband Digital Predistortion Using Spectral Extrapolation of Band-Limited Feedback Signal,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, 61 (2014), 2088–2097

[48] Bai Dongwoon, Cheolhee Park, Jungwon Lee, Hoang Nguyen, Jaspreet Singh, Ankit Gupta, and others, “LTE-Advanced Modem Design: Challenges and Perspectives,” *Communications Magazine, IEEE*, 50 (2012), 178–186

[49] Iwamura Mikio, Kamran Etemad, Mo-Han Fong, Ravi Nory, and Robert Love, “Carrier Aggregation Framework in 3GPP LTE-Advanced [WiMAX/LTE Update],” *Communications Magazine, IEEE*, 48 (2010), 60–67

[50] Park Chester Sungchung, A Khayrallah, and others, “Carrier Aggregation for LTE-Advanced: Design Challenges of Terminals,” *Communications Magazine, IEEE*, 51 (2013), 76–84

[51] Pedersen Klaus Ingemann, Frank Frederiksen, Claudio Rosa, Hung Nguyen, Luis Guilherme Uzeda Garcia, and Yuanye Wang, “Carrier Aggregation for LTE-Advanced: Functionality and Performance Aspects,” *Communications Magazine, IEEE*, 49 (2011), 89–95

[52] Baschirotto Andrea, R Castello, F Campi, Giovanni Cesura, M Toma, R Guerrieri, and others, “Baseband Analog Front-End and Digital Back-End for Reconfigurable Multi-Standard Terminals,” *Circuits and Systems Magazine, IEEE*, 6 (2006), 8–28

[53] Li Ushan, Steve McLaughlin, David G M Cruickshank, and others, “Towards Multi-Mode Terminals,” *Vehicular Technology Magazine, IEEE*, 1 (2006), 17–24

[54] Hu Rose Qingyang, and Yi Qian, “An Energy Efficient and Spectrum Efficient Wireless Heterogeneous Network Framework for 5G Systems,” *Communications Magazine, IEEE*, 52 (2014), 94–101

[55] Lehman Thomas, Jerry Sobieski, and Bijan Jabbari, “DRAGON: a Framework for Service Provisioning in Heterogeneous Grid Networks,” *Communications Magazine, IEEE*, 44 (2006), 84–90

[56] Tzanakaki Anna, Markos P Anastasopoulos, Georgios S Zervas, Bijan Rahimzadeh Rofoee, Reza Nejabati, and Dimitra Simeonidou, "Virtualization of Heterogeneous Wireless-Optical Network and IT Infrastructures in Support of Cloud and Mobile Cloud Services," *Communications Magazine, IEEE*, 51 (2013)

[57] Chen Yan, Shunqing Zhang, Shugong Xu, and Geoffrey Ye Li, "Fundamental Trade-Offs on Green Wireless Networks," *Communications Magazine, IEEE*, 49 (2011), 30–37

[58] Han Congzheng, Tim Harrold, Simon Armour, Ioannis Krikidis, Stefan Videv, Peter M Grant, and others, "Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks," *Communications Magazine, IEEE*, 49 (2011), 46–54

[59] Tranter William, K Shanmugan, Theodore Rappaport, and Kurt Kosbar, *Principles of Communication Systems Simulation with Wireless Applications* (Prentice Hall Press, 2003)

[60] Swarts Francis, Pieter Rooyen, Ian Oppermann, and Michiel P Lötter, CDMA Techniques for Third Generation Mobile Systems (Springer, 1999)

[61] Tachikawa Keiji, W-CDMA Mobile Communications System (John Wiley & Sons, 2003)

[62] Dahlman Erik, Stefan Parkvall, and Johan Skold, 4G: LTE/LTE-Advanced for Mobile Broadband (Academic Press, 2013)

[63] Dahlman Erik, Stefan Parkvall, Johan Skold, and Per Beming, 3G Evolution: HSPA and LTE for Mobile Broadband (Academic press, 2010)

[64] Hazewinkel Michiel, "Central limit theorem", Encyclopedia of Mathematics, Springer, ISBN 978-1-55608-010-4 (2001)

[65] Han Seung Hee, and Jae Hong Lee, "An Overview of Peak-to-Average Power Ratio Reduction Techniques for Multicarrier Transmission," *Wireless Communications, IEEE*, 12 (2005), 56–65

[66] Jiang Tao, and Yiyan Wu, "An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals," *IEEE transactions on broadcasting*, 54 (2008), 257–268

[67] Wunder Gerhard, Robert F H Fischer, Holger Boche, Simon Litsyn, and Jong-Seon No, "The PAPR Problem in OFDM Transmission: New Directions for a Long-Lasting Problem," *arXiv preprint arXiv:1212.2865*, 2012

[68] "LogiCORE IP peak cancellation crest factor reduction v3.0 datasheet," Xilinx Inc., San Jose, CA, USA, Jun. 2011.

[69] "Crest Factor Reduction Processor," Texas Instruments Inc., TX, USA, Apr. 2009.

[70] "Global mobile statistics 2014 Part A: Mobile subscribers; handset market share; mobile operators," available online: http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a

[71] Saleh Adel A M, "Frequency-Independent and Frequency-Dependent Nonlinear Models of TWT Amplifiers," *Communications, IEEE Transactions on*, 29 (1981), 1715–1720

[72] Isaksson Magnus, David Wisell, and Daniel Ronnow, "A Comparative Analysis of Behavioral Models for RF Power Amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, 54 (2006), 348–359

[73] Muhonen Kathleen J, Mohsen Kavehrad, and Rajeev Krishnamoorthy, "Look-up Table Techniques for Adaptive Digital Predistortion: a Development and Comparison," *Vehicular Technology, IEEE Transactions on*, 49 (2000), 1995–2002

[74] Pedro José C, and Stephen A Maas, "A Comparative Overview of Microwave and Wireless Power-Amplifier Behavioral Modeling Approaches," *Microwave Theory and techniques, IEEE Transactions on*, 53 (2005), 1150–1163

[75] Zhu Anding, and Thomas J Brazil, "An Overview of Volterra Series Based Behavioral Modeling of RF/microwave Power Amplifiers," in *Proc. Wireless Microw. Technol. Conf*, 2006, pp. 1–5

[76] Caverly Robert, Frederick Raab, and Joseph Staudinger, "High-Efficiency Power Amplifiers," *Microwave Magazine, IEEE*, 13 (2012), S22–S32

[77] Nemati Hossein Mashad, Paul Saad, Christian Fager, and Kristoffer Andersson, "High-Efficiency Power Amplifier," *Microwave Magazine, IEEE*, 12 (2011), 81–84

[78] Shimbo Osamu, "Effects of Intermodulation, AM-PM Conversion, and Additive Noise in Multicarrier TWT Systems," *Proceedings of the IEEE*, 59 (1971), 230–238

[79] Ku Hyunchul, Michael D McKinley, and J Stevenson Kenney, “Quantifying Memory Effects in RF Power Amplifiers,” *Microwave Theory and Techniques, IEEE Transactions on*, 50 (2002), 2843–2849

[80] Vuolevi Joel H K, Timo Rahkonen, and Jani P A Manninen, “Measurement Technique for Characterizing Memory Effects in RF Power Amplifiers,” *Microwave Theory and Techniques, IEEE Transactions on*, 49 (2001), 1383–1389

[81] Bosch Wolfgang, and Giuliano Gatti, “Measurement and Simulation of Memory Effects in Predistortion Linearizers,” *Microwave Theory and Techniques, IEEE Transactions on*, 37 (1989), 1885–1890

[82] Aikio Janne P, and Timo Rahkonen, “A Comprehensive Analysis of AM--AM and AM--PM Conversion in an LDMOS RF Power Amplifier,” *Microwave Theory and Techniques, IEEE Transactions on*, 57 (2009), 262–270

[83] Boumaiza Slim, and Fadhel M Ghannouchi, “Thermal Memory Effects Modeling and Compensation in RF Power Amplifiers and Predistortion Linearizers,” *Microwave Theory and Techniques, IEEE Transactions on*, 51 (2003), 2427–2433

[84] Vuolevi Joel, and Timo Rahkonen, *Distortion in RF Power Amplifiers* (Artech house, 2003)

[85] Ding Lei, and Guo Tong Zhou, “Effects of Even-Order Nonlinear Terms on Power Amplifier Modeling and Predistortion Linearization,” *Vehicular Technology, IEEE Transactions on*, 53 (2004), 156–162

[86] Raich Raviv, Hua Qian, and G Tong Zhou, “Orthogonal Polynomials for Power Amplifier Modeling and Predistorter Design,” *Vehicular Technology, IEEE Transactions on*, 53 (2004), 1468–1479

[87] “Givens Rotation”, Wikipedia, 2014, online available: http://en.wikipedia.org/wiki/Givens_rotation

[88] “Gram–Schmidt process” Wikipedia, 2014, online available: http://en.wikipedia.org/wiki/Gram%E2%80%93Schmidt_process

[89] Björck Ake, “Solving Linear Least Squares Problems by Gram-Schmidt Orthogonalization,” *BIT Numerical Mathematics*, 7 (1967), 1–21

[90] Crespo-Cadenas Carlos, Javier Reina-Tosina, María J Madero-Ayora, and Jesús Muñoz-Cruzado, “A New Approach to Pruning Volterra Models for Power Amplifiers,” *Signal Processing, IEEE Transactions on*, 58 (2010), 2113–2120

[91] Golub Gene H, and Charles F Van Loan, *Matrix Computations* (JHU Press, 2012)

[92] Schetzen Martin, “Theory of Pth-Order Inverses of Nonlinear Systems,” *Circuits and Systems, IEEE Transactions on*, 23 (1976), 285–291

[93] Sarti Augusto, and Silvano Pupolin, “Recursive Techniques for the Synthesis of a Pth-Order Inverse of a Volterra System,” *European Transactions on Telecommunications*, 3 (1992), 315–322

[94] Tsimbinos John, and Kenneth V Lever, “Nonlinear System Compensation Based on Orthogonal Polynomial Inverses,” *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 48 (2001), 406–417

[95] Ma Yuelin and Yasushi Yamao, “A Method for Inverting Monotonous Continuous Nonlinear Function,” 2014 TriSAI Workshop (2014), pp. 1-5.

[96] Ma Yuelin, Yoshihiko Akaiwa, and Yasushi Yamao, “Fast Baseband Polynomial Inverse Algorithm for Nonlinear System Compensation,” in *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, 2012, pp. 1–5

[97] Yuelin MA, Yasushi Yamao, and Yoshihiko Akaiwa, “An Algorithm for Obtaining the Inverse for a Given Polynomial in Baseband,” *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, 96 (2013), 675–683

[98] Bussgang Julian Jakob, “Crosscorrelation Functions of Amplitude-Distorted Gaussian Signals”, 1952

[99] Ma Yuelin, Yoshihiko Akaiwa, Yasushi Yamao, and Songbai He, “Test Bed for Characterization and Predistortion of Power Amplifiers,” *International Journal of RF and Microwave Computer-Aided Engineering*, 23 (2013), 74–82

[100] Abidi Asad A, “Direct-Conversion Radio Transceivers for Digital Communications,” *Solid-State Circuits, IEEE Journal of*, 30 (1995), 1399–1410

[101] Cavers James K, and Maria W Liao, “Adaptive Compensation for Imbalance and Offset Losses in Direct Conversion Transceivers,” *Vehicular Technology, IEEE Transactions on*, 42 (1993), 581–588

[102] Smith Steven W, and others, “The Scientist and Engineer’s Guide to Digital Signal Processing”, 1997

[103] Abbas Muhammad, Oscar Gustafsson, and H åkan Johansson, “On the Fixed-Point Implementation of Fractional-Delay Filters Based on the Farrow Structure,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 60 (2013), 926–937

[104] Lee Wei Rong, Lou Caccetta, and Volker Rehbock, “Optimal Design of All-Pass Variable Fractional-Delay Digital Filters,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 55 (2008), 1248–1256

[105] Valimaki Vesa, and Timo I Laakso, “Principles of Fractional Delay Filters,” in *Acoustics, Speech, and Signal Processing, 2000. ICASSP’00. Proceedings. 2000 IEEE International Conference on*, 2000, VI, 3870–3873

[106] Yekta M M Jahani, “Wideband Maximally Flat Fractional-Delay Allpass Filters,” *Electronics Letters*, 46 (2010), 722–723

[107] Zhu Anding, Paul J Draxler, Jonmei J Yan, Thomas J Brazil, Donald F Kimball, and Peter M Asbeck, “Open-Loop Digital Predistorter for RF Power Amplifiers Using Dynamic Deviation Reduction-Based Volterra Series,” *Microwave Theory and Techniques, IEEE Transactions on*, 56 (2008), 1524–1534

[108] Ghannouchi Fadhel M, and Oualid Hammi, “Behavioral Modeling and Predistortion,” *Microwave Magazine, IEEE*, 10 (2009), 52–64

[109] Hammi Oualid, and Fadhel M Ghannouchi, “Power Alignment of Digital Predistorters for Power< Newline/> Amplifiers Linearity Optimization,” *Broadcasting, IEEE Transactions on*, 55 (2009), 109–114

[110] Gilabert Pere L, Albert Cesari, Gabriel Montoro, Eduard Bertran, and J-M Dilhac, “Multi-Lookup Table FPGA Implementation of an Adaptive Digital Predistorter for Linearizing RF Power Amplifiers with Memory Effects,” *Microwave Theory and Techniques, IEEE Transactions on*, 56 (2008), 372–384

[111] Presti Calogero D, Donald F Kimball, and Peter M Asbeck, “Closed-Loop Digital Predistortion System with Fast Real-Time Adaptation Applied to a Handset WCDMA PA Module,” *Microwave Theory and Techniques, IEEE Transactions on*, 60 (2012), 604–618

[112] Eweda Eweda, “Transient Performance Degradation of the LMS, RLS, Sign, Signed Regressor, and Sign-Sign Algorithms with Data Correlation,” *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, 46 (1999), 1055–1062

[113] Cavers James K, “Optimum Table Spacing in Predistorting Amplifier Linearizers,” *Vehicular Technology, IEEE Transactions on*, 48 (1999), 1699–1705

[114] “$\mu$-law algorithm”, online available: http://en.wikipedia.org/wiki/%CE%9C-law_algorithm

[115] XILINX Inc., LogiCORE IP CORDIC v4.0 datasheet, Mar., 2011.

[116] “Alpha max plus beta min algorithm,” online available: http://en.wikipedia.org/wiki/Alpha_max_plus_beta_min_algorithm

[117] Smith Steven W, and others, “The Scientist and Engineer’s Guide to Digital Signal Processing”, 1997

[118] Freescale Semiconductor Inc., MDE6IC9120NR1 datasheet, Nov. 2009.

[119] 3GPP TS 36.300, “Evolved Universal Terrestrial Radio Access (E-UTRA); Stage 2 Description,” Version 9.3.0, Apr. 2010.

[120] Iwamatsu T, Jianmin Zhou, Hui Li, Zhan Shi, Pei Chen, Bingbing Zhao, and others, “A Simple Predistorter in Fixed Microwave Radio System with the Transmission Power Control,” in *Power Amplifiers for Wireless and Radio Applications (PAWR), 2012 IEEE Topical Conference on*, 2012, pp. 65–68

[121] Koeppl Heinz, and Peter Singerl, “An Efficient Scheme for Nonlinear Modeling and Predistortion in Mixed-Signal Systems,” *Circuits and Systems II: Express Briefs, IEEE Transactions on*, 53 (2006), 1368–1372

[122] Yu Chao, Lei Guan, Erni Zhu, and Anding Zhu, “Band-Limited Volterra Series-Based Digital Predistortion for Wideband RF Power Amplifiers,” *Microwave Theory and Techniques, IEEE Transactions on*, 60 (2012), 4198–4208

[123] Yu Xin, “Digital Predistortion Using Feedback Signal with Incomplete Spectral Information,” in *Microwave Conference Proceedings (APMC), 2012 Asia-Pacific*, 2012, pp. 950–952

[124] Oppenheim Alan V, Ronald W Schafer, John R Buck, and others, *Discrete-Time Signal Processing* (Prentice-hall Englewood Cliffs, 1989), II

[125] Jain A, and Surendra Ranganath, “Extrapolation Algorithms for Discrete Signals with Application in Spectral Estimation,” *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29 (1981), 830–845

[126] Papoulis Athanasios, “A New Algorithm in Spectral Analysis and Band-Limited Extrapolation,” *Circuits and Systems, IEEE Transactions on*, 22 (1975), 735–742

[127] Kelley Carl T, *Iterative Methods for Optimization* (Siam, 1999), XVIII

[128] Dogancay, Kutluyil, “Blind Compensation of Nonlinear Distortion for Bandlimited Signals,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 52 (2005), 1872–1882

[129] Stapleton Shawn P, and Flaviu C Costescu, “An Adaptive Predistorter for a Power Amplifier Based on Adjacent Channel Emissions [mobile Communications],” *Vehicular Technology, IEEE Transactions on*, 41 (1992), 49–56

[130] Pei Soo-Chang, and Jong-Jy Shyu, “Design of Complex FIR Filters with Arbitrary Complex Frequency Responses by Two Real Chebyshev Approximations,” *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on*, 44 (1997), 170–174

[131] Hashemi Hossein, and Ali Hajimiri, “Concurrent Multiband Low-Noise Amplifiers--Theory, Design, and Applications,” *IEEE Transactions on Microwave Theory and techniques*, 50 (2002), 288–301

[132] V. Steel and A. Morris, “Tunable RF technology overview,” *Microwave Journal.* vol. 55, No. 11, pp. 26-31, Nov. 2012.

[133] Zou Qiyue, Mohyee Mikhemar, and Ali H Sayed, “Digital Compensation of Cross-Modulation Distortion in Software-Defined Radios,” *Selected Topics in Signal Processing, IEEE Journal of*, 3 (2009), 348–361

[134] Habibi Hooman, Erwin J G Janssen, W Yan, and Jan W M Bergmans, “Digital Compensation of Cross-Modulation Distortion in Multimode Transceivers,” *IET communications*, 6 (2012), 1724–1733

[135] Bassam Seyed Aidin, Mohamed Helaoui, and Fadhel M Ghannouchi, “2-D Digital Predistortion (2-D-DPD) Architecture for Concurrent Dual-Band Transmitters,” *Microwave Theory and Techniques, IEEE Transactions on*, 59 (2011), 2547–2553

[136] “HMC374/374E SMT PHEMT low noise amplifier datasheet”, Hittite Microwave, Chelmsford, MA, USA, 2009.

[137] Bassam S, Wenhua Chen, Mohamed Helaoui, and F Ghannouchi, “Transmitter Architecture for CA: Carrier Aggregation in LTE-Advanced Systems,” *Microwave Magazine, IEEE*, 14 (2013), 78–86

[138] Rawat Karun, Mohammad S Hashmi, and Fadhel M Ghannouchi, “Dual-Band RF Circuits and Components for Multi-Standard Software Defined Radios,” *Circuits and Systems Magazine, IEEE*, 12 (2012), 12–32

[139] Roblin Patrick, Christophe Quindroit, Naveen Naraharisetti, Shahin Gheitanchi, and Mike Fitton, “Concurrent Linearization: The State of the Art for Modeling and Linearization of Multiband Power Amplifiers,” *Microwave Magazine, IEEE*, 14 (2013), 75–91

[140] Bassam Seyed Aidin, Andrew Kwan, Wenhua Chen, Mohamed Helaoui, and Fadhel M Ghannouchi, “Subsampling Feedback Loop Applicable to Concurrent Dual-Band Linearization Architecture,” *Microwave Theory and Techniques, IEEE Transactions on*, 60 (2012), 1990–1999

[141] Cidronali A, I Magrini, R Fagotti, and G Manes, “A New Approach for Concurrent Dual-Band IF Digital Predistortion: System Design and Analysis,” in *Integrated Nonlinear Microwave and Millimetre-Wave Circuits, 2008. INMMIC 2008. Workshop on*, 2008, pp. 127–130

[142] Nader Charles, Wendy Van Moer, Kurt Barbé, Niclas Bjorsell, and Peter Handel, “Harmonic Sampling and Reconstruction of Wideband Undersampled Waveforms: Breaking the Code,” *Microwave Theory and Techniques, IEEE Transactions on*, 59 (2011), 2961–2969

# Publications

## *Journal Papers*

1. Yuelin Ma, Yasushi Yamao, Yoshihiko Akaiwa, Koji Ishibashi, “Wideband digital predistortion using spectral extrapolation of band-limited feedback signal”, *IEEE Transactions on Circuits and Systems I*, vol.61, no.7, pp.2088-2097, Jul. 2014. (relate to Chapter 6)
2. Yuelin Ma, Yasushi Yamao, Koji Ishibashi and Yoshihiko Akaiwa, "Adaptive compensation of inter-band modulation distortion for tunable concurrent dual-band receivers," *IEEE Transactions on Microwave Theory and Techniques*, vol.61, no.12, pp.4209-4219, Dec. 2013. (relate to Chapter 8)
3. Yuelin Ma, Yasushi Yamao, Yoshihiko Akaiwa and Chunlei Yu, "FPGA implementation of adaptive digital predistorter with fast convergence rate and low complexity for multi-channel transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol.61, no.11, pp.3961-3973, Nov. 2013. (relate to Chapter 5)
4. Yuelin Ma, Yasushi Yamao and Yoshihiko Akaiwa, “An algorithm for obtaining the inverse for a given polynomial in baseband,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E96-A, no. 3, pp. 675–683, Mar. 2013. (relate to Chapter 3)
5. Yuelin Ma, Yasushi Yamao, Yoshihiko Akaiwa and Songbai He, “Test bed for characterization and predistortion of power amplifiers,” *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 23, no. 1, pp. 74–82, Jan. 2013. (relate to Chapter 4)
6. Yuelin Ma, Songbai He, Yoshihiko Akaiwa and Yasushi Yamao “An open-loop digital predistorter based on memory polynomial inverses for linearization of RF power amplifier,” *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 21, no. 5, pp. 589–595, Sep. 2011. (relate to Chapter 3)
7. Yuelin Ma, Yasushi Yamao, “Spectra-Folding Feedback Architecture for Concurrent Dual-Band Power Amplifier predistortion”, submitted to *IEEE Transactions on Microwave Theory and Techniques.* (relate to Chapter 9)

## *International Conference Papers*

8. Yuelin Ma and Yasushi Yamao, “A Method for Inverting Monotonous Continuous Nonlinear Function,” 2014 TriSAI Workshop, pp. 1-5. Bejing, China, Sep. 2014. (relate to Chapter 3)
9. Yuelin Ma, Yasushi Yamao and Yoshihiko Akaiwa, “Blind Nonlinear Compensation Technique for RF Receiver Front-End,” 43rd European Microwave Conference, pp. 1527-1530. Nuremburg, Germany, Oct. 2013. (relate to Chapter 7)
10. Yuelin Ma, Yasushi Yamao and Yoshihiko Akaiwa, “Fast Baseband Polynomial Inverse Algorithm for Nonlinear System Compensation,” IEEE Proceeding of Vehicular Technology Conference 2012-Spring, 2F-1, Yokohama, Japan, May 2012. (relate to Chapter 3)
11. Yuelin Ma, Yoshihiko Akaiwa, Yasushi Yamao, “A Novel Feedback Configured Predistorter for Radio Power Amplifiers with Memory Effect,” IEEE VTS Proceeding of Asia-

Pacific Wireless Communications Symposium, S12-2, Kaohsiung, Taiwan, May 2010. (relate to Chapter 3)

12. Yuelin Ma and Yasushi Yamao, “Experimental Results of Digital Predistorter for Very Wideband Mobile Communication System,” IEEE Proceeding of Vehicular Technology Conference 2015-Spring, Glasgow, Scotland, May 2015. (relate to Chapter 6)

*Domestic Conference Papers*

13. Yuelin Ma and Yasushi Yamao「Power Amplifier Linearization Using Digital Predistortion Technology: Practical Development and Implementation Issues」， 信学技報, vol. 114, no. 498, MW2014-213, pp. 61-66, 2015 年 3 月. (relate to Chapter 5)
14. 馬岳林，山尾泰「受信フロントエンドの非線形ひずみに対するブラインド補償法」， 信学技報, vol. 113, no. 379, MW2013-190, pp. 85-90, 2014 年 1 月. (relate to Chapter 7)
15. 馬岳林，山尾泰，赤岩芳彦「入力信号帯域幅に制限された情報を基に補償を行うプリディストーション非線形補償の効果」， 信学技報, vol. 112, no. 443, RCS2012-352, pp. 405-410, 2013 年 2 月. (relate to Chapter 6)
16. Yuelin Ma, Yoshihiko Akaiwa and Yasushi Yamao「A Novel Feedback Configured Predistorter for Power Amplifiers with Memory Effect」， 信学技報, vol. 110, no. 127, RCS2010-50, pp. 13-17, 2010 年 7 月. (relate to Chapter 3)
17. 船橋鴻志，馬岳林，山尾泰「零点を有する2次Δ-Σ変調器を用いたEPWM送信回路のFPGA設計」，2014信学総大，A-1-14，2014年3月