# Study on complexity reduction of digital predistortion for power amplifier linearization

Siqi Wang

UNIVERSITÉ PARIS-EST

ÉCOLE DOCTORALE MATHÉMATIQUES, SCIENCES, ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

DOCTORAL THESIS

IN ELECTRONICS, OPTRONICS AND SYSTEM

*Presented by*

Siqi WANG

# Study on Complexity Reduction of Behavioral Modeling Digital Predistortion for Power Amplifier Linearization

*Supervised by*

Geneviève BAUDOIN, Olivier VENARD, Mazen ABI HUSSEIN

December 30, 2017

**Jury:**

| | | |
|---|---|---|
| Rapporteur | **Jean-François HELARD** | INSA de Rennes |
| Rapporteur | **Smaïl BACHIR** | XLIM |
| Examinateur | **Yves LOUET** | CENTRALSUPELEC, Rennes |
| Examinateur | **Sylvain TRAVERSO** | THALES Communication |
| Examinateur | **Daniel ROVIRAS** | CNAM |
| Examinateur | **Myriam ARIAUDO** | ENSEA |
| Directeur de thèse | **Geneviève BAUDOIN** | ESIEE Paris |
| co-Directeur de thèse | **Olivier VENARD** | ESIEE Paris |

Version Provisoire

# Acknowledgement

I would like to express my gratitude to the people without whom this dissertation would not have been completed.

First, I would like to thank my supervisor, Pr. Geneviève Baudoin and my co-supervisors Olivier Venard and Mazen Abi Hussein who continuously supported my work and ideas throughout my three years in Université Paris-Est and ESIEE Paris. We had very frequent fruitful discussions since the first day of my PhD life. The progress of my study was prominent thanks to their precious advices and kind patience. It was a great joy to have been working with them.

Next, I would like to thank my thesis defense committee: Dr. Jean-François HELARD, Dr. Smaïl Bachir, Pr. Daniel ROVIRAS, Dr. Yves Louet, Dr. Myriam Ariaudo and Dr. Sylvain Traverso for helping to improve my dissertation with their thoughtful advice and valuable suggestions.

I would like to thank all members of the laboratory ESYCOM for the enthusiastic ambiance, amiable encouragement and convivial company during three years.

Last but not least, I would like to thank my family for their unrequited love and priceless support .

# Abstract

This dissertation contributes to the linearization techniques of high power amplifier using digital predistortion method. High power amplifier is one of the most nonlinear components in radio transmitters. Baseband adaptive digital predistortion is a powerful technique to linearize the power amplifiers and allows to push the power amplifier operation point towards its high efficiency region. Linearization of power amplifiers using digital predistortion with low complexities is the focus of this dissertation. An algorithm is proposed to determine an optimal model structure of single-stage or multi-stage predistorter according to a trade-off between modeling accuracy and model complexity. Multi-stage cascaded digital predistortions are studied with different identification methods, which have advantages on complexity of model identification compared with single-stage structure. In terms of experimental implementations, this dissertation studies the impact of different gain choices on linearized power amplifier. All studies are evaluated with a Doherty power amplifier.



**Keywords: Power amplifier Digital predistortion Nonlinear distortion Cascaded model Optimization**

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| RF | Radio Frequency |
| PA | Power Amplifier |
| PD | Predistortion |
| DPD | Digital Predistortion |
| ACPR | Adjacent Channel Power Ratio |
| NMSE | Normalized Mean Square Error |
| EVM | Error Vector Magnitude |
| PAPR | Peak-to-Average Power Ratio |
| MP | Memory Polynomial |
| GMP | Generalized Memory Polynomial |
| DDR | Dynamic Deviation Reduction |
| OFDM | Orthogonal Frequency Division Multiplex |
| LTE | Long Term Evolution |
| AWG | Arbitrary Waveform Generator |
| VSA | Vector Signal Analyzer |
| PAE | Power Added Efficiency |
| OPBO | Output Peak Back-Off |
| PDF | Probability Density Function |
| HC | Hill Climbing |
| FCHC | First-Choice Hill Climbing |
| GA | Genetic Algorithms |

# Publications

**Journals**:
1. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "A Novel Algorithm for Determining the Structure of Digital Predistortion Models", submitted to *IEEE Transactions on Vehicular Technology* in December 2016. First review in July 2017. Resubmitted in November 2017.

**International conferences**:
1. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Optimal Sizing of Generalized Memory Polynomial Model Structure Based on Hill-Climbing Heuristic", *European Microwave Conference (EuMC) 2016* in London.
2. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Comparison of hill-climbing and genetic algorithms for digital predistortion models sizing", *IEEE International Conference on Electronics, Circuits and Systems (ICECS) 2016* in Monaco.
3. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Impact of the Normalization Gain of Digital Predistortion on Linearization Performance and Power Added Efficiency of the Linearized Power Amplifier", *European Microwave Conference (EuMC) 2017* in Nuremberg.
4. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Performance Analysis of Multi-stage Cascaded Digital Predistortion", *IEEE International Conference on Telecommunications and Signal Processing (TSP) 2017* in Barcelona.
5. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Optimal Sizing of Cascaded Digital Predistortion for Linearization of High Power Amplifiers", *2017 IEEE Asia Pacific Microwave Conference (APMC2017)* in Kuala Lumpur.
6. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Identification of Low Order Cascaded Digital Predistortion with Different-structure

Stages for Linearization of Power Amplifiers", *2018 IEEE Radio & Wireless Week (RWW)* in Anaheim.

**National conference**:
1. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "L'algorithme "first choice hill climbing" pour le dimensionnement du modèle polynomial à mémoire généralisé", *Journée Nationales Micro-Ondes (JNM) 2017*.

**Participations to GDR Soc-Sip**:
1. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "Optimal Sizing of Generalized Memory Polynomial Model Structure Based on Hill-Climbing Heuristic", *GdR SoC-SiP Colloque 2016*.
2. Siqi Wang, Mazen Abi Hussein, Olivier Venard and Genevieve Baudoin, "L'algorithme "first choice hill climbing" pour le dimensionnement du modèle polynomial à mémoire généralisé", *GdR SoC-SiP Colloque 2017*.

# Chapter 1

# Introduction

## 1.1 Motivation and Objective

High power amplifier (PA) is one of the most nonlinear components in radio transmitters. It is a critical element of radio transmitters in current and future generations of wireless systems and is responsible for a large amount of the power consumption. So the power autonomy and the size of transmitters strongly depend on the power efficiency of the PA. Unfortunately, for most current types of PA, e.g. the Doherty Power Amplifier [1], a good efficiency is obtained at the price of a poor linearity specially with modern communication waveforms that have very high peak-to-average power ratio (PAPR) and large bandwidths. High efficiency and linearity are two important requirements which are not easy to fulfill simultaneously. For high efficiency, PAs are usually driven towards saturation region where high nonlinear behavior is exhibited. PAs may have not only very strong nonlinearities but also memory effects [2]. Baseband adaptive digital predistortion (DPD) is a powerful technique to linearize the PA and allows to push the PA operation point towards its high efficiency region. The principle of DPD is to apply, upstream of PA, a pre-correction on the signal so that the cascade of the DPD and the PA is close to an ideal linear, memoryless system. In this way, the PA can be driven more towards the high efficiency saturation region without compromising much on linearity.

Different models for DPD have been proposed during the recent decades of years. The first objective of this dissertation is to present how to compensate for the nonlinearities and memory effects of PA and improve its efficiency using DPD method. Different DPDs are analyzed experimentally. The mathematical model

of DPD should have low complexity so that its implementation on hardware is viable.

Reducing the model complexity and the complexity of DPD implementation is always a key topic. Multi-stage cascaded model has been studied in [3]. The identification complexity is able to be reduced by decomposing a single-stage model into multi-stage of simpler models. The multi-stage cascaded model is very different from a single-stage model. Different identification orders result in different lienarization performances.

In another side, the dimension of the DPD models directly decides the model complexity. Good modeling accuracy demands large dimension. However, it is not always true in reverse. A method which can determine the structure of a DPD with a good trade-off between modeling accuracy and model complexity is widely needed. As the performances of DPD models depend on the PA characteristics and its input signal, this method should be applied once the PA or input signal is changed. In this case, the execution time of this method needs to be short.

This dissertation discusses the techniques of DPD that have been studied and makes new contributions on optimal DPD model structure determination and cascaded DPDs. The optimal cascaded DPD structrue can be also determined using the proposed method.

This dissertation focuses mainly on:

- Searching the optimal DPD mathematical model structure according to a given criterion which represents the trade-off between modeling accuracy and model complexity.
- Identification algorithm and performance of cascaded DPD models, and comparison between multi-stage cascaded models and single stage models.
- Experimental evaluation of the proposed techniques using a high power. A high power Doherty PA is tested with different linearization methods.

## 1.2 Main Contributions

The contributions of this dissertation are listed as follows:

- An algorithm based on hill-climbing heuristic is proposed to determine the optimal structure of DPD model according to a given criterion. Its effectiveness has been confirmed in case of generalized memory polynomial models.

  - Different criteria are proposed to represent the trade-off between modeling accuracy and model complexity.
  - Different methods to accelerate the algorithm are proposed and studied.

- One-stage DPD and multi-stage cascaded DPD model are studied and applied on test bench.
  - Different methods to identify the cascaded DPD are compared and discussed.
  - An efficient way to determine the structure of cascaded DPD models is also proposed and confirmed on test bench.

- An adjustment of implementation in measurement of PA with linear gain is proposed. The experimental linearizations of PA in this dissertation are based on it. The impact of different choices of gain is also studied.

## 1.3 Outline

The dissertation is organized as follows.

Chapter 2 gives general concepts and the background of PA and DPD. Different DPD mathematical models are cited and reviewed in this chapter. The test bench for experimental implementations is also introduced.

The algorithm based on hill-climbing heuristic to determine the optimal structure of a generalized memory polynomial (GMP) model is described in Chapter 3. Different optimizations of the algorithm are proposed and studied and the solutions of GMP model structures are tested on test bench with a Three-way Doherty PA. This algorithm is compared with genetic algorithm (GA).

In Chapter 4, cascaded DPD is introduced to linearize PA with less complexity of identification. The cascaded DPD is proved to have less coefficient dynamic range and lower conditioning number of matrix while potentially keeping similar linearization performance compared with one-stage DPD. Different identification methods of cascaded DPD are studied and compared. The optimal structure of cascaded DPD can be determined by the algorithm proposed in Chapter 3.

Chapter 5 shows the experimental results on test bench. The test bench calibration algorithm is investigated. The characterization of PA is implemented in this chapter and different gains for linearization are chosen and compared.

Finally Chapter 6 gives the conclusion and prospects.

# Chapter 2

# Generalities on Power Amplifiers and Linearization Techniques

## 2.1 Introduction

The nonlinearity and the efficiency of PA depend on the input signal amplitude as shown in Fig 2.1. The blue curve is the output power of PA in function of input power, and the red curve is the efficiency of the PA in function of input power. To get the maximal efficiency, it is better to move the operating point near the saturation zone. However, the signal falls into the nonlinear zone and undesired distortion comes out. To avoid nonlinear spectral distortion, the operating point needs to be backed off away from the saturation zone.

There are three linearization techniques to compensate for the distortion of PA: the feedforward technique, the feedback loop, and the predistortion (PD). The disadvantages of feedforwad approach are its high hardware complexity, limitations on operating point of PA and on efficiency limits. The feedback loop has limitation on bandwidth of the signal.

There have been numerous studies on implementations of predistortion: analog predistortion [4] [5] [6] [7] and digital predistortion [8] [9] [10] [11]. The former is implemented on analog hardware using nonlinear components, which limits its performance [9]. The latter has better adaptability and performance for signals of bandwidths up to several tens of MHz [12]. For ultra wide bandwidth, signals as generated by carrier aggregation, analog predistortion may be a better solution [13]. In this dissertation, we discuss only about digital predistortion.

This chapter is organized as follows. Section 2 discusses the nonlinearity of

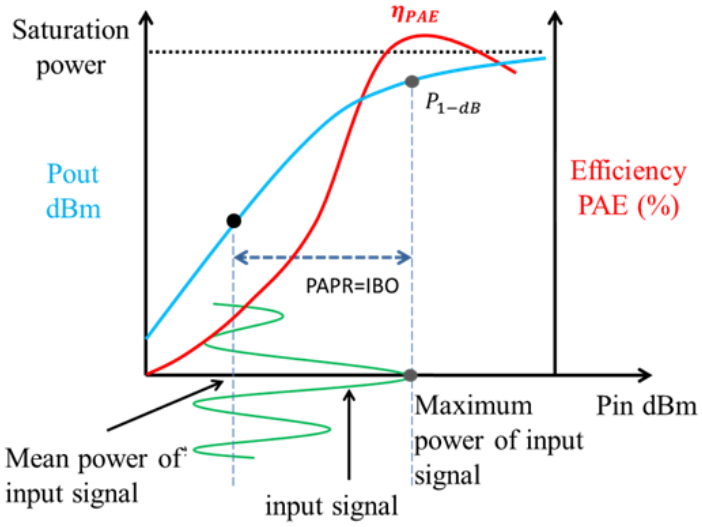



Figure 2.1: Trade-off between the linearity and the efficiency of PA

PA. Section 3 defines some criteria to evaluate the PA effects on signals. In Section 4, main concepts of DPD are introduced. Different models for PA and DPD modeling are presented in Section 5. Section 6 introduces the techniques used to identify the coefficients of models. The test bench is presented in Section 7. The conclusion is given in Section 8.

## 2.2 Distortions Introduced by Power Amplifiers

### 2.2.1 Nonlinearity

The nonlinearity of PA can be shown with characteristic curves which are called AM/AM & AM/PM (Amplitude Modulation/Amplitude Modulation & Amplitude Modulation/Phase Modulation) curves as shown in Fig 2.2. The blue curve is AM/AM curve which shows the power of a Three-way Doherty PA's output signal magnitude in function of its input signal magnitude. The orange curve is AM/PM curve which shows the phase deviation of this PA's output signal in function of its input signal magnitude at the fundamental frequency $f_0$.

We can see that the gain is compressed by PA when the input power increases. The 1dB compression point ($P_{1dB}$) is defined at the point where the compression

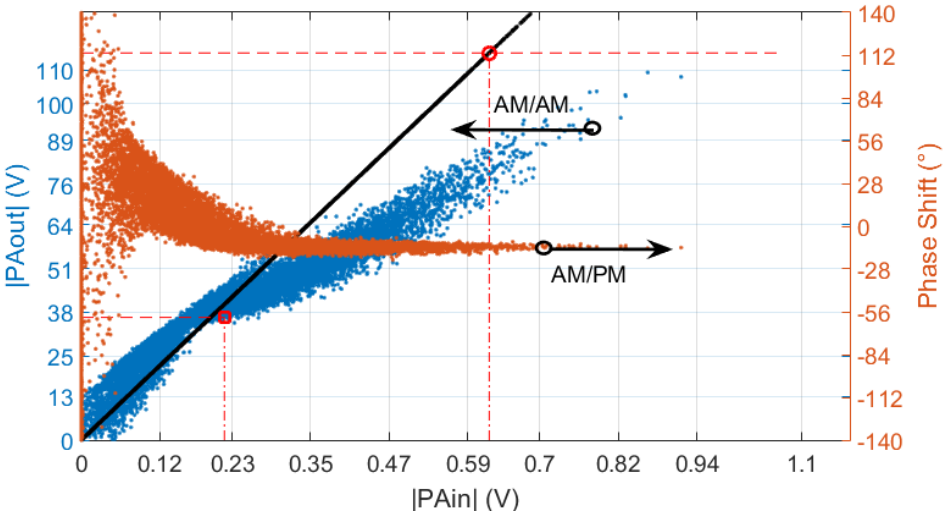



Figure 2.2: AM/AM & AM/PM curve of a Three-way Doherty PA

of gain equals to 1 dB. The 3dB compression point ($P_{3dB}$) is defined at the point where the compression of gain equals to 3 dB.

To avoid saturation at the PA output, we need to keep the PA input peak power within a threshold as shown in Fig 2.3. The PA output peak power is denoted by $P_{Peak}$. An output peak back-off (OPBO) is expressed as:

$$OPBO_{dB} = P_{Sat} - P_{Peak} \tag{2.1}$$

where $P_{Sat}$ is the saturated output power of PA.

### 2.2.2 Harmonics and Intermodulation Products

The nonlinearity of PA can be approximated with a power series:

$$y(t) = \sum_{k=1}^{K_a} a_k x^k(t) \tag{2.2}$$

where $x(t)$ and $y(t)$ are the input and output signal of PA, $K_a$ is the highest order of nonlinearity.

When the incident signal is a single tone signal as

$$x(t) = Acos(2\pi f_0 t), \tag{2.3}$$

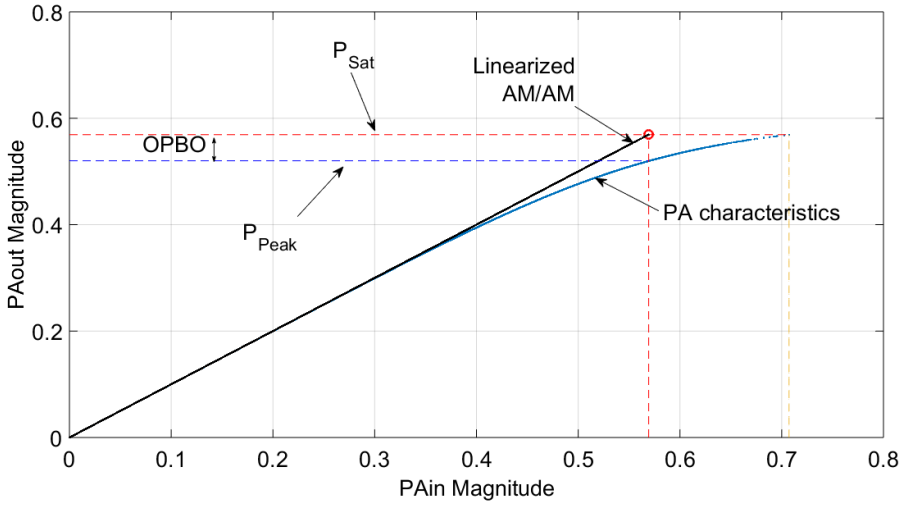



Figure 2.3: Application of OPBO

the output signal can be written as

$$y(t) = \sum_{k=1}^{K_a} b_k cos(2\pi k f_0 t). \tag{2.4}$$

The terms corresponding to $k > 1$, which are multiples of the original frequency, are called the harmonics. Fig 2.4 shows the harmonics generated by a Three-way Doherty PA excited by an 2 GHz one-tone signal. The spectrum shows the spikes at 4 GHz and 6 GHz frequencies.

When the incident signal is a two-tone signal as

$$x(t) = A_1 cos(2\pi f_1 t) + A_2 cos(2\pi f_2 t) \tag{2.5}$$

where $|f_1 - f_2|$ is very small compared with the carrier frequency in transmission, there are more spectral components generated at the output of PA. The frequencies of these components can be expressed as:

$$f_{com} = pf_1 + qf_2. \tag{2.6}$$

The order of the term is decided by $N = |p| + |q|$. These components are called intermodulation (IMD) products. The most important distortion generally results

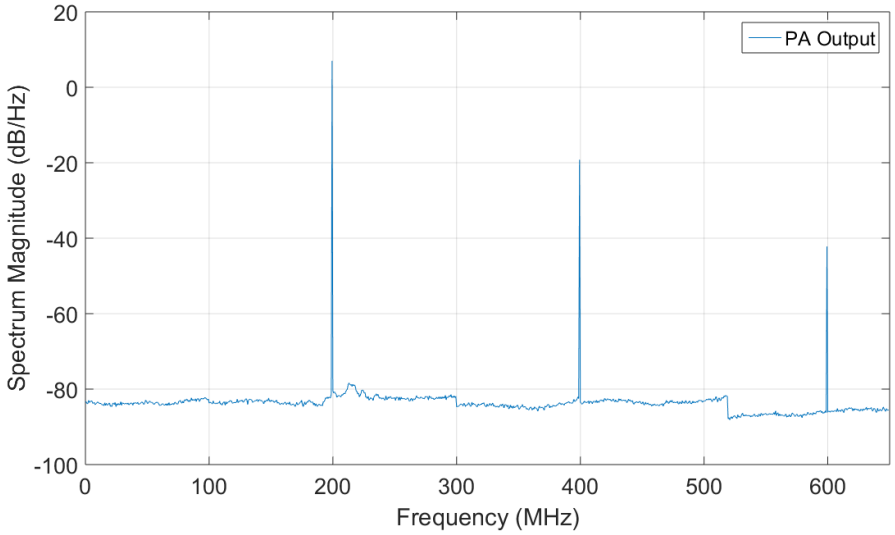



Figure 2.4: Harmonics at PA output excited by a 2 GHz one-tone signal

from the third order IMD products (IMD3) nearest to $f_1$ and $f_2$: at frequencies of $2f_1 - f_2$ and $2f_2 - f_1$. A 1 & 6 MHz two-tone signal is used to study the IMD products. The signal is modulated by a carrier of 2.14 GHz and is sent to a Three-way Doherty PA. Fig 2.5 shows the IMD generated by the PA after the PA output signal is demodulated to baseband. In this case we can see IMD3, IMD5, IMD7 and IMD9. We can also observe some adjoint spikes beside each intermodulation spike which are resulted from the direct component ($f = 0$) introduced by the equipment.

### 2.2.3 Memory Effect

The AM/AM & AM/PM curves in Fig 2.6 represent the characteristics of a memoryless PA model. However, the AM/AM & AM/PM curves of most of high power amplifiers exhibit strong dispersions as shown in Fig 2.2. This is caused by memory effects [2]. The output signal of a PA depends on both the present and the historical input signal [12]. Thus one power of input signal may correspond to several different powers of output signal.

The memory effects can be categorized in terms of their time constant compared with the reciprocal of their bandwidth: long-term memory effects with large time constant, and short-term memory effects with low time constant. Large

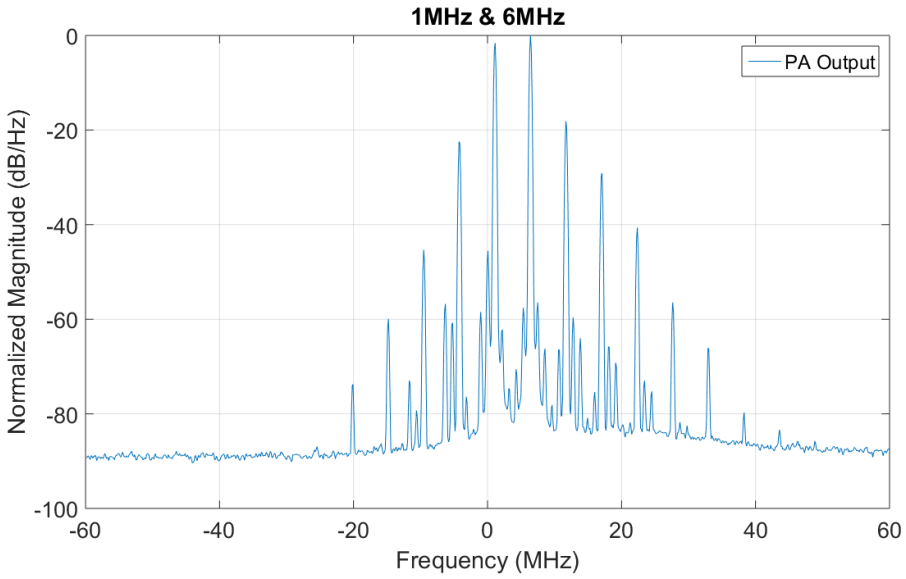


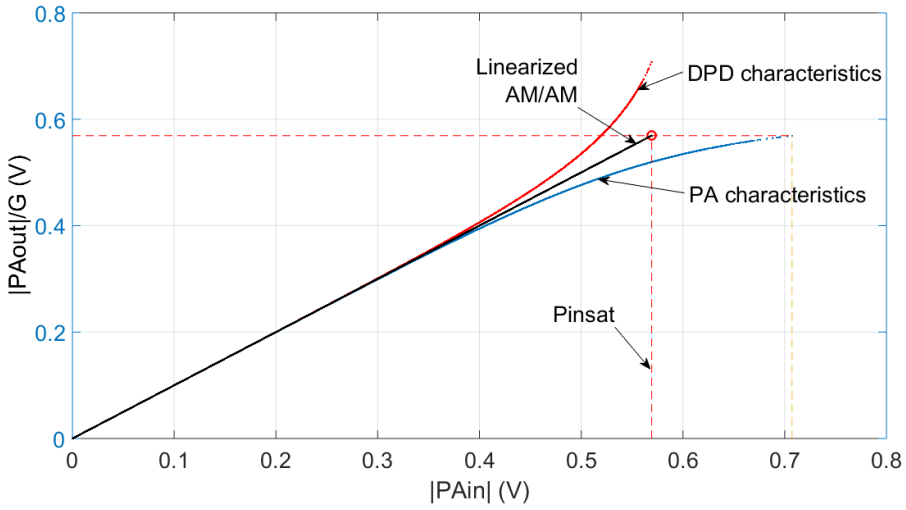

Figure 2.5: Intermodulation products at PA output excited by a 1 & 6 MHz two-tone signal

time-constant memory effects are mainly due to thermal effects and biasing circuits [14]. Short-time constant memory effects are due to short time constants of biasing circuits.

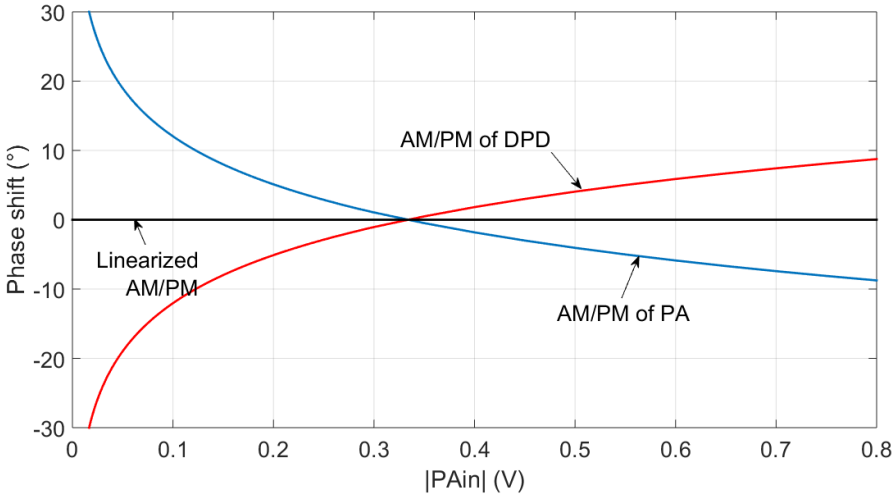
There are different sources of memory effects [12].

The electrical memory effects are caused by the variation of circuit component (transistors, matching networks and bias networks) impedances in function of different signal modulation frequencies [15]. For the wide-bandwidth signal, the impedances variations of varying envelope can be very large, which is the main source of memory effects compared with that of fundamental or second harmonic impedances. Since the envelope frequency range covers from dc to the maximum modulation frequency, it produces both long-term and short-term memory effects. Besides, the modulation will generate a dc component in harmonics which introduces the bias voltage variation.

The thermal memory effects [16] are caused by electro-thermal couplings. The dissipated power of transistors increases the temperature which may affect the electrical parameters of the transistors. Since the temperature is not changed instantaneously, the thermal memory effects are long-term memory effects.




(a) AM/AM curve




(b) AM/PM curve

Figure 2.6: PA linearization with DPD

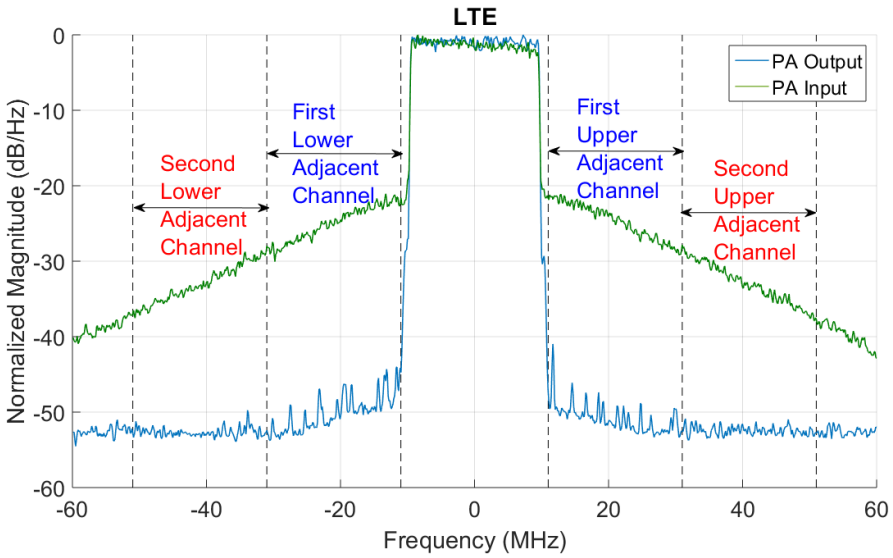



Figure 2.7: Upper and Lower Adjacent Channels

## 2.3 Parameters to Evaluate PA Effects on Signals

The distortion introduced by PA could be evaluated with the normalized mean square error (NMSE), the adjacent channel power ratio (ACPR) and Error Vector Magnitude (EVM) [17].

The NMSE between the PA output signal $y(n)$ and the desired PA output signal $\tilde{y}(n)$ (proportional to the input signal $x(n)$), expressed as

$$\text{NMSE}_{\text{dB}} = 10log_{10}\left[\frac{\sum_{n=1}^{N}|\tilde{y}(n)-y(n)|^2}{\sum_{n=1}^{N}|\tilde{y}(n)|^2}\right], \tag{2.7}$$

is used to evaluate both in-band and out-band distortion of PA.

The ACPR at PA output is used to evaluate the out-of-band distortion of PA:

$$\text{ACPR}_{\text{dB}} = 10log_{10}\left[\frac{\sum_{\omega\in L}|Y(\omega)|^2+\sum_{\omega\in U}|Y(\omega)|^2}{\sum_{\omega\in M}|Y(\omega)|^2}\right] \tag{2.8}$$

where $L$ and $U$ are the first lower adjacent channel frequencies and the first upper adjacent channel frequencies as shown in Fig 2.7, respectively, $M$ is the main channel frequencies, $Y(\omega)$ is Discrete Fourier Transform of $y(n)$.

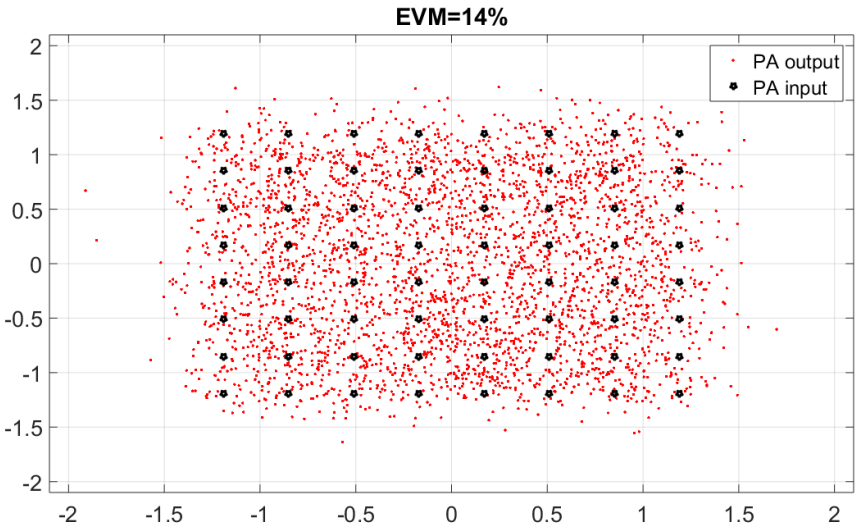



Figure 2.8: Constellation of a 64 QAM signal with EVM=14%

The error vector magnitude (EVM) is used to evaluate the in-band distortion of PA. It is applied for the constellation of modulated signals.

$$\text{EVM}_{\%} = \sqrt{\frac{\frac{1}{N}\sum_{j=0}^{N-1}(\delta I^2 + \delta Q^2)}{S_{avg}^2}} \times 100\% \tag{2.9}$$

where $\delta I$ and $\delta Q$ are errors magnitude corresponding to in-phase symbol and quadrature symbol of received data compared with an ideally reconstructed constellation respectively, $N$ is the number of symbols, $S_{avg}^2$ is the average square magnitude of $N$ symbols. The offset and the rotation of the constellation can be also taken into consideration in the definition of EVM. Fig 2.8 shows the constellation of a 64 QAM signal. The black stars are ideally reconstructed constellation. The red points are received data. The EVM in this case is 14%.

## 2.4 Principle of Digital Predistortion

Baseband adaptive digital predistortion (DPD) is a powerful technique to compensate for nonlinearities and memory effects of the PA. Theoretically a DPD has

inverse characteristics of that of the PA. In Fig 2.6, the blue curve is the AM/AM curve of PA, which has a nonlinearity when the input signal power approaches $P_{InSat}$, the saturation input power. The PD's characteristics can be obtained by reversing that of PA, as the red curve. By combining PD and PA, the AM/AM curve of the entire system become linear (at least up to a maximum limit) as the black curve shows.

The DPD is required to have high linearization performances and low cost. The coefficients of DPD can be identified in different ways which are thoroughly introduced in Section 2.6. The cost of DPD can be the model complexity of DPD or the complexity of its identification procedure, which is influenced by the number of coefficients of the DPD model.

## 2.5 Different PA and DPD Models

Behavioral modeling of PA considers the PA as a black-box and only its input signal and output signal are needed. The DPD design and identification can be independent of deep knowledge of the RF circuit physics and functionality [18].

Numerous mathematical models have been proposed to model PAs and to serve as DPD. Most of them are based on Volterra Series Model. A commonly used model which can compensate for both nonlinearities and memory effects is Memory Polynomial (MP) model [8] [19]. However, MP model may have limited performance when the PA exhibits strong nonlinearities and memory effects.

In recent years there has been growing interest in more complex models derived from Volterra series. In MP model which is also called the diagonal Volterra model [8], only the diagonal terms are used, and all off-diagonal terms are zero. A "near-diagonality" structure is proposed in [20] and it has been shown that the off-diagonal terms may be more important than the diagonal terms. The Generalized Memory Polynomial (GMP) model [21], Laguerre-Volterra model [22], Kautz-Volterra model [23] and dynamic-deviation-reduction (DDR) Volterra model [24] have been proposed using different pruning techniques of the Volterra series. These models are linear combinations of some basis functions.

According to the input signal, we may have RF model or baseband model [25]. In case of RF model, the input signal is modulated with a carrier frequency which is much greater than the bandwidth of the signal. The baseband model is used to make a study of the envelope of the RF signal, which is equivalent to a signal centered to zero frequency.

### 2.5.1 Memoryless and quasi-memoryless Models

The DPD models can be divided into two categories: memoryless models and models with memory. For low power amplifier or narrow-bandwidth input signal, the characteristics of PA can be modeled as a memoryless or quasi-memoryless model, e.g. polynomial model, Saleh model [26] and Rapp model [27]. A memoryless model describes only AM/AM conversion.

We use a baseband equivalent model to study the RF system because it requires lower sampling frequency compared with the carrier frequency. The baseband signal $x(n)$ is complex signal. Considering that $x(n-l) \approx x(n)$, the relation between PA input and output signal can be presented as a memoryless polynomial model:

$$y(n) = \sum_{k=0}^{K-1} b_k |x(n)|^k x(n), \tag{2.10}$$

where $x(n)$ and $y(n)$ are the baseband PA input and output signal respectively, and $b_k$ is the complex-valued coefficient of $k$th order of nonlinearity, $K$ is the maximum order of nonlinearity. However, in this case, the model can compensate for the phase shift and it creates also an AM/PM conversion. Thus the model is called quasi-memoryless model.

For quasi-memoryless models, there are many different models. Here we present Saleh model and Rapp model which are two of the most used quasi-memoryless models. Saleh model is one of the first proposed to model a TWT amplifier with two-parameter formulas [26]:

$$\begin{aligned} A(x(n)) &= \frac{\alpha_A |x(n)|}{1+\beta_A |x(n)|^2} \\ \phi(x(n)) &= \frac{\alpha_\phi |x(n)|^2}{1+\beta_\phi |x(n)|^2} \end{aligned} \tag{2.11}$$

where $\alpha_A$, $\beta_A$, $\alpha_\phi$ and $\beta_\phi$ are constants, $A(x(n))$ and $\phi(x(n))$ represent AM/AM and AM/PM curves respectively.

Rapp model is proposed in [27] by replacing the formula of $A(x(n))$ with:

$$A(x(n)) = \frac{G|x(n)|}{(1+(\frac{G|x(n)|}{A_0})^{2p})^{\frac{1}{2p}}} \tag{2.12}$$

where $A_0$ is the maximum output power of PA, $G$ is the small signal gain, $p$ is used to control the smoothness of the AM/AM curve near saturation zone.

### 2.5.2 Models Derived from Volterra Series

The electrical cause and electro-thermal couplings introduce memory effects which can seriously limit the performance of DPD. For high power amplifiers with wide band applications, the memory effects are too strong to be neglected. In this case, the Volterra series model is a good choice to take into account these memory effects.

The Volterra series model for radio frequency (RF) system can be expressed as:

$$\tilde{y}(n) = \sum_{k=1}^{+\infty} \int_{0}^{+\infty} \cdots \int_{0}^{+\infty} h_k(\tau_1, .., \tau_k) \prod_{j=1}^{k} \tilde{x}(t-\tau_j) d\tau_j \tag{2.13}$$

where $\tilde{x}$ is the RF input signal, $h_k(\cdot)$ is the real-valued $k$-th order Volterra kernel.

After demodulating the RF signal to baseband, the envelope of signal can be obtained by a low-pass filter. Thus we can have baseband equivalent discrete time Volterra series model with complex signal $x(n)$ input:

$$y(n) = \sum_{k=0}^{K} \sum_{l_1=0}^{L-1} \sum_{l_2=l_1}^{L-1} \cdots \sum_{l_{2k+1}=l_{2k}}^{L-1} h_{2k+1}(l_1, .., l_{2k+1}) \prod_{j=1}^{k+1} x(n-l_j) \prod_{j=k+2}^{2k+1} x^*(n-l_j) \tag{2.14}$$

where $L$ represents the memory depth and $K$ represents the order of nonlinearity.

The modeling performance of Volterra series model strongly depends on the number of terms. The large number of terms make the full Volterra series model very complicated and time consuming to identify.

#### Memory Polynomial Model and Generalized Memory Polynomial Model

Memory polynomial (MP) model is a particular case of Volterra series model, which has only the diagonal terms. Suppose $x(n)$ is the baseband input signal and $y(n)$ is the output signal of PD, the MP model is represented as (2.15):

$$y(n) = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} c_{kl} x(n-l)|x(n-l)|^k. \tag{2.15}$$

Though the MP model is proven effective in predistortion modeling for nonlinear PA with memory effect, we can still achieve even better performance by formulating more general memory structures. As the basis functions in MP model are

diagonal in terms of memory, the off-diagonal terms $x(n-l)|x(n-m)|^k$, where $l \neq m$, are added to form the generalized memory polynomial model (GMP).

The GMP model is widely used [28] [29] [30] [31], and it has been shown in [32] that it has a good trade-off for accuracy versus complexity. The GMP can be written as:

$$\begin{aligned} y(n) = & \sum_{k=0}^{\mathcal{K}_a-1} \sum_{l=0}^{\mathcal{L}_a-1} a_{kl} x(n-l)|x(n-l)|^k \\ & + \sum_{k=1}^{\mathcal{K}_b} \sum_{l=0}^{\mathcal{L}_b-1} \sum_{m=1}^{\mathcal{M}_b} b_{klm} x(n-l)|x(n-l-m)|^k \\ & + \sum_{k=1}^{\mathcal{K}_c} \sum_{l=0}^{\mathcal{L}_c-1} \sum_{m=1}^{\mathcal{M}_c} c_{klm} x(n-l)|x(n-l+m)|^k \end{aligned} \tag{2.16}$$

where the DPD input is $x(n)$, the DPD output is $y(n)$, $k$ is the index for nonlinearity, and $l$, $m$ are the indices for memory. $a_{kl}$, $b_{klm}$, $c_{klm}$ are the complex coefficients of the signal and envelope, the signal and lagging envelope, and the signal and leading envelope, respectively. $\mathcal{K}_a$, $\mathcal{K}_b$, $\mathcal{K}_c$ are the highest orders of nonlinearity. $\mathcal{L}_a$, $\mathcal{L}_b$, $\mathcal{L}_c$ are the highest memory depths. $\mathcal{M}_b$, $\mathcal{M}_c$ denote the longest lagging and leading delay tap length, respectively.

The GMP model may outperform MP model on reducing spectral regrowth or adjacent channel power ratio (ACPR) by adding more model complexity. But as the number of terms increasing, the model structure sizing and its coefficients identification become more complicated [21].

### Orthogonal Polynomial Model

The basis functions of conventional polynomial models as (2.15) are not orthogonal. The polynomial model with orthogonal basis proposed in [33] alleviates the numerical instability problem associated wtih the conventional polynomials and generally yield better modeling accuracy.

The conventional memoryless polynomial model (2.10) is written in a new way with orthogonal basis $\psi_k(|x|)$

$$y(n) = \sum_{k=0}^{K-1} \beta_k \psi_k(x(n)) \tag{2.17}$$

where

$$\psi_k(x) = \sum_{i=0}^{k} U_{ik}|x|^i x \tag{2.18}$$

and $U_{ik}$ are the coefficients of orthogonal polynomial basis functions.

The orthogonal polynomials depend on the probability density functions of signal amplitude. When the absolute module of input complex signal $|x|$ is uniformly distributed in $[0, 1]$, we can have:

$$U_{lk} = \begin{cases} (-1)^{i+k}\frac{(k+i)!}{(i-1)!(i+1)!(k-i)!}, & i \leqslant k \\ 0, & i > k \end{cases} \tag{2.19}$$

The transform of memory polynomial model is also proposed [33]:

$$y[n] = \sum_{k=0}^{K-1}\sum_{l=0}^{L-1} \beta_{kl}\psi_{kl}(x(n-l)). \tag{2.20}$$

In this case, the orthogonality is kept only among the terms of the same delay.

### 2.5.3 Block-Oriented Nonlinear Systems

The nonlinearities and memory effects can be modeled separately by the association of linear time invariant (LTI) dynamic blocks and static nonlinear blocks: Block-oriented nonlinear (BONL) system [34].

Hammerstein, Wiener, and Wiener-Hammerstein models are widely used BONL systems and their identification algorithms are thoroughly researched. Suppose $u(n)$ is the input signal and $x(n)$ is the output signal of the predistorter. If we model the nonlinear part with a polynomial, the Hammerstein model is given in (2.21) and (2.22):

$$w(n) = \sum_{k=0}^{K} a_k u(n)|u(n)|^k \tag{2.21}$$

$$x(n) = \sum_{l=0}^{L} b_l w(n-l). \tag{2.22}$$

where $w(n)$ is the intermediate signal between the 2 stages of the model.

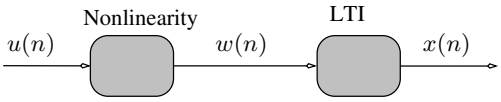



Figure 2.9: Hammerstein model

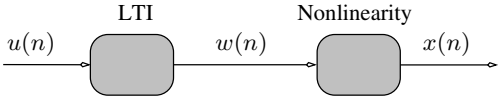



Figure 2.10: Wiener model

In Hammerstein model, it represents the output of Nonlinear part and the input of LTI part as shown in Fig 2.9. In Wiener model, it represents the output of LTI part and the input of Nonlinear part as shown in Fig 2.10.

The memoryless nonlinearity part is represented by (2.21). The LTI system is represented by (2.22). And the Wiener model is given in (2.23) and (2.24):

$$w(n) = \sum_{l=0}^{L} a_l u(n-l) \tag{2.23}$$

$$x(n) = \sum_{k=0}^{K} b_k w(n)|w(n)|^k. \tag{2.24}$$

The LTI system is represented by (2.23) and the memoryless nonlinearity part is represented by (2.24).

The Wiener-Hammerstein model is given in (2.25), (2.26) and (2.27):

$$w(n) = \sum_{l=0}^{L} a_l u(n-l) \tag{2.25}$$

$$s(n) = \sum_{k=0}^{K} b_k w(n)|w(n)|^k \tag{2.26}$$

$$x(n) = \sum_{m=0}^{M} c_m s(n-m). \tag{2.27}$$

It is a memoryless nonlinearity sandwiched between two linear filters, as shown in Fig 2.11. Wiener model and Hammerstein model can be considered as the particular cases of Wiener-Hammerstein model [35].

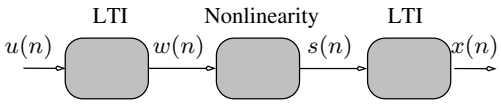
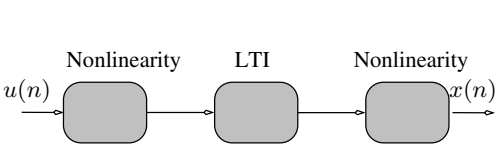
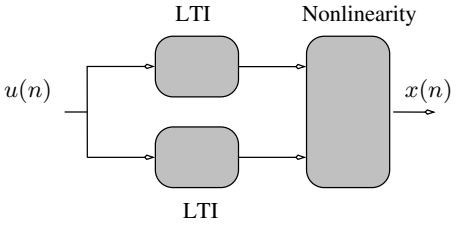


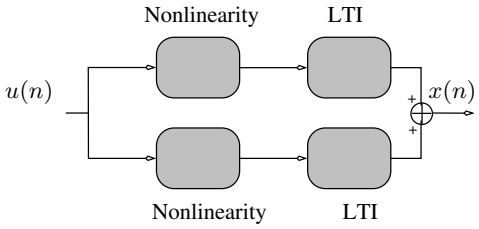

Figure 2.11: Wiener-Hammerstein model



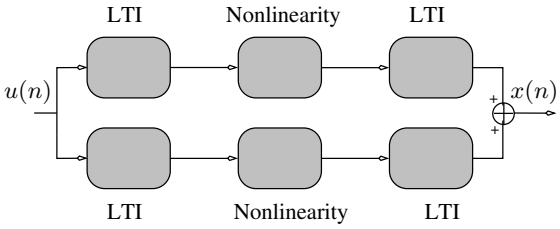

(a) Hammerstein-Wiener model




(b) Parallel Wiener model




(c) Parallel Hammerstein model




(d) Parallel Wiener-Hammerstein model

Figure 2.12: Different BONL systems

With different combinations, we can derive also Hammerstein-Wiener model, Parallel Wiener model, Parallel Hammerstein model, Parallel Wiener-Hammerstein model as shown in Fig 2.12 [36].

### 2.5.4 Polynomial Model with Separable Functions

A general structure of DPD is described in [37] using separable functions.

For any nonlinear model with memory depth $L$, the relation between input $x$ and output $y$ can be defined with an operator $P : C^L \rightarrow C$ that

$$y(n) = P(x(n), x(n-1), ..., x(n-L+1)). \tag{2.28}$$

We denote the function of PA or DPD model by $P_0$. As $x(n)^{k+1}$ can be decomposed into the product $x(n)|x(n)|^k$ when $k$ is even, an approximation is made

from a multivariate function to a sum of separable functions:

$$\begin{aligned} P_0(x(n), x(n-1), ..., x(n-L+1)) &\approx \sum_{k=1}^{K} \prod_{l=0}^{L-1} P_{kl}(x(n-1)) \\ &= \sum_{k=1}^{K} x(n-m_k) \prod_{l=0}^{L-1} P_{kl}(|x(n-l)|) \end{aligned} \tag{2.29}$$

where $0 \leqslant m_k \leqslant L-1$. The separation of term $x(n-m_k)$ allows to replace the function of complex variable $P_{kl}(x(n-l))$ by the function of real variable $P_{kl}(|x(n-l)|)$, which reduces the complexity of computation.

The systematic structure (2.29) is such a general structure that many DPD models proposed in literature are some particular cases of it [37]. The basis functions $P_{kl}$ can be polynomials, sinusoidal functions and LUTs, or mixture of them.

The separable functions can be represented also by polynomials, which are expressed in terms of a group of orthonormal basis $\{\psi_0(|x|), ..., \psi_{M-1}(|x|)\}$:

$$P_{kl}(|x|) = \sum_{m=0}^{M-1} u_m^{kl} \psi_m(|x|) \tag{2.30}$$

where $\psi_m(|x|)$ is an orthogonal polynomial of degree $m$ whose weight function is the probability density function (PDF) of the transmitted signal. The coefficients $u_m^{kl}$ can be estimated by solving a system of linear equations.

### 2.5.5 Vector-switched Model and Decomposed Vector Rotation Model

Instead of modeling the PA or DPD with one model, we can also construct several submodels according to different amplitudes of input signal.

The DPD using a vector-switched model is proposed in [38]. It applies Volterra series or GMP models with different nonlinearity orders according to the amplitudes of the input signal segments. The partitioning of these segments needs reasonable decision borders, which can be solved as a vector quantization problem [39].

The amplitude space is divided into several switching regions which are designed based on the current and previous complex-valued input samples of the training set. Every region has a centroid. With the amplitudes of input samples

of the training set, the closest centroid could be found and then the corresponding switching region is selected. This technique can linearize PAs with irregular input/output characteristics at low computational complexity. The potential concern is that when different models are used for consecutive samples, there will be some distortion introduced by the switching.

A Decomposed Vector Rotation (DVR) model which is different from Volterra models is proposed in [40]. It is based on the canonical piecewise-linear function (CPWL) [41] which works in the discrete time domain and can represent both static nonlinearity and memory effects. The DVR model inherits these features and furthermore satisfies linear-in-parameters condition of DPD model selection and dealing with complex-valued signals. The representation of DVR model is written below:

$$\begin{aligned} y(n) = & \sum_{i=0}^{M} a_i x(n-i) \\ & + \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,1} ||x(n-i)| - \beta_k| e^{j\theta(n-i)} \\ & + \sum_{k=1}^{K} \sum_{i=0}^{M} c_{ki,21} ||x(n-i)| - \beta_k| e^{j\theta(n-i)} \cdot |x(n)| \\ & + \sum_{k=1}^{K} \sum_{i=1}^{M} c_{ki,22} ||x(n-i)| - \beta_k| \cdot x(n) \\ & + \sum_{k=1}^{K} \sum_{i=1}^{M} c_{ki,23} ||x(n-i)| - \beta_k| \cdot x(n-i) \\ & + \sum_{k=1}^{K} \sum_{i=1}^{M} c_{ki,24} ||x(n)| - \beta_k| \cdot x(n-i) \\ & + \cdots \end{aligned} \tag{2.31}$$

where $x(n)$ and $y(n)$ are input and output of the model respectively, $\beta_k$ is the breakpoint, $K$ is the number of breakpoints, $M$ is the memory depth.

Very high order of nonlinearities can be characterized by this model with a small number of terms. And it is much more flexible and capable in modeling highly nonlinear and "unusual" PAs compared to the Volterra models [40]. This

representation (2.31) is simplified in [42] by rewriting the 1st-order basis

$$\sum_{k=1}^{K}\sum_{i=0}^{M} c_{ki,1}||x(n-i)|-\beta_k|e^{j\theta(n-i)} \tag{2.32}$$

as a summation

$$\sum_{i=0}^{M}\delta_{ki}x(n-i)+\sum_{i=0}^{M}x(n-i)\sum_{k=1}^{K}\alpha_{ki}||x(n-i)-\beta_k| \tag{2.33}$$

assuming the CPWL is transfered back to the polynomials. The first basis of the second order

$$\sum_{k=1}^{K}\sum_{i=0}^{M} c_{ki,21}||x(n-i)|-\beta_k|e^{j\theta(n-i)}\cdot|x(n)| \tag{2.34}$$

is also replaced by

$$\sum_{k=1}^{K}\sum_{i=0}^{M} c_{ki,21}||x(n-i)|-\beta_k|x(n-i)\cdot|x(n)| \tag{2.35}$$

The calculation of exponential functions are thus approximated by calculating polynomials, which reduces the model complexity while keeping nearly the same modeling performance.

### 2.5.6 Neural Network Models

Neural network (NN) is another choice for PA and DPD modeling. Multilayer perceptron NN is well used because it can be trained to learn any arbitrary nonlinear input-output relationships from corresponding data [43]. As developed from imitating the biological nervous system, a multilayer perceptron neural network consists of an input layer, some hidden layers, and an output layer. The input signal is fed to the input layer, and the output signal is found at the output layer. Each layer is a group of neurons which have no connection between each other but have connections with the neurons of the next layer.

Multilayer perceptron neural network is illustrated in Fig 2.13. The total number of layers is $L$, where $L \geqslant 3$. Each layer has different number of neurons. The memory depth is represented by the number of samples $N_1$ at the input layer.

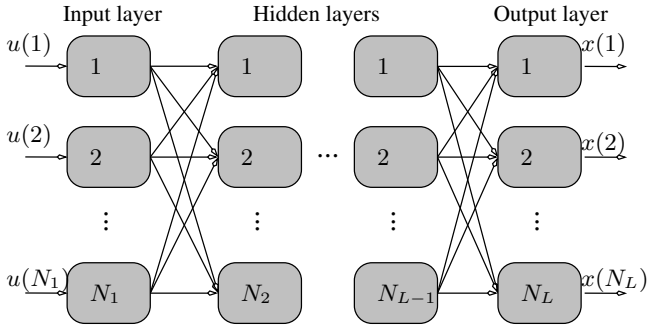



Figure 2.13: Multilayer Perceptron Neural Network model

Different artificial neural network models have been applied on PA and DPD modeling. A real-valued time delay neural network is proposed in [44] using only real-valued parameters and the real components of input and output signals. Thus the complexity is largely reduced with complex modulated signals having highly time-varying envelopes. In [45], a distributed spatiotemporal neural network based model is proposed to model the PA/transmitter with modulated signals. It is proved to have low computational load and fast convergence.

## 2.6 DPD Model Identification

An important aspect of digital predistortion is the estimation of the digital predistorter model coefficients.

Two approaches can be used to estimate model coefficients: indirect learning architecture (ILA) and direct learning architecture (DLA) [46]. In the ILA approach, a post-inverse block of the PA is first identified with the input and output signals of the PA and then applied upstream of PA as a DPD. In the DLA approach, the DPD is directly identified with the input and output signals of the system [47].

In the following, we consider the case where the model is linear with respect to its coefficients such as Volterra series, MP, GMP, DDR models and many other models derived from Volterra series.

In this section, we discuss only the identification of single block DPD. The identification in the case of multi-blocks DPD such as BONL or cascade DPD is introduced in Chapter 4.

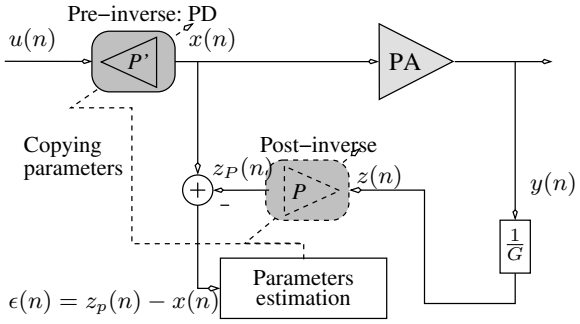



Figure 2.14: Indirect Learning Architecture

### 2.6.1 Indirect Learning Architecture

The indirect learning architecture (ILA), which is depicted in Fig 2.14, is a simple architecture to identify DPD models using different techniques such as least squares (LS), least mean square (LMS) or recursive least square (RLS) [19] [48] [46] [3]. Here we describe the LS approach.

A post-inverse of the PA is identified and used as a DPD. The aim is to minimize LS criterion built on the difference between the output $z_p$ of the postdistorter and the input $x$ of the PA. The instantaneous error is defined as $\epsilon(n) = z_p(n) - x(n)$.

For the post-inverse block of models linear with respect to their coefficients, the relation between its input and output can be rewritten using matrix notation for a block of $N$ samples:

$$\boldsymbol{z_p} = \mathbf{Z}\boldsymbol{c} \tag{2.36}$$

where $\boldsymbol{z_p}$= $[z_p(1), \ldots, z_p(N)]^T$, $\boldsymbol{z}$= $[z(1), \ldots, z(N)]^T$, $\boldsymbol{c}$ is a $R \times 1$ vector containing the set of coefficients $c_{kl}$, $\mathbf{Z}$ is $N \times R$ matrix containing basis functions of $\boldsymbol{z}$. For example, in the case of a memory polynomial model, this matrix is represented as:

$$\mathbf{Z} = \begin{bmatrix} \Phi_{1,1}(z(n)) & \cdots & \Phi_{K,1}(z(n)) & \Phi_{1,2}(z(n)) & \cdots & \Phi_{K,L}(z(n)) \\ \Phi_{1,1}(z(n-1)) & \cdots & \cdot & \cdot & \cdots & \Phi_{K,L}(z(n-1)) \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \Phi_{1,1}(z(n-N+1)) & \cdots & \cdot & \cdot & \cdots & \cdot \end{bmatrix}, \tag{2.37}$$

where $\Phi_{k,l}(z(n)) = z(n-l+1)|z(n-l+1)|^{k-1}$ and $R = KL$ the total number of coefficients of an MP model.

Table 2.1: Computational complexity of each step in DPD identification

| Step | Number of flops |
|---|---|
| 1 | $2NR^2 - \frac{2}{3}R^3$ |
| 2 | $2NR - R$ |
| 3 | $R^2$ |

The least square (LS) solution will be the solution for the following equation

$$[\mathbf{Z}^H\mathbf{Z}]\hat{\boldsymbol{c}} = \mathbf{Z}^H\boldsymbol{x} \tag{2.38}$$

which minimizes the LS cost function

$$C = \sum_{n=1}^{N} |z_p(n) - x(n)|^2. \tag{2.39}$$

Many approaches can be used to solve (2.38). It should be noticed that the matrix $\mathbf{Z}^H\mathbf{Z}$ is generally badly conditioned. A possible technique is to use QR decomposition. Eq (2.38) can be solved by three steps:

- In step 1, we compute a QR factorization $\mathbf{Z} = \mathbf{QR}$, where $\mathbf{Q}$ is a $N \times N$ square matrix and $\mathbf{R}$ is a $N \times R$ upper-triangle matrix. In case of Householder triangularization method, there are $2N \times R^2 - \frac{2}{3}R^3$ flops (floating point operations) [49].
- In step 2, the matrix $\mathbf{Q}^H\boldsymbol{x}$ is computed. There are $2NR - R$ flops in this step.
- In step 3, an upper-triangle matrix $\mathbf{R}\hat{\boldsymbol{c}} = \mathbf{Q}^H\boldsymbol{x}$ is solved for $\hat{\boldsymbol{c}}$. There are $R^2$ flops in this step.

Table 2.1 summarizes roughly the number of complex multiplications needed in the post-inverse identification, where $N$ is the length of dataset for DPD identification. If N is large compared with $R$, we can estimate the computation load by $\mathcal{O}(2NR^2)$.

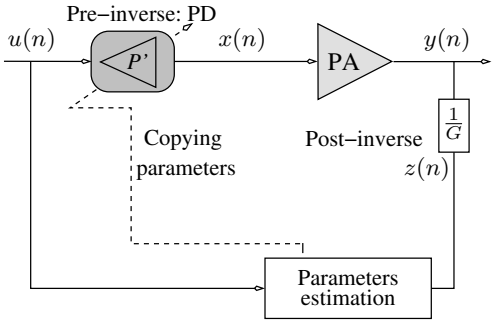



Figure 2.15: Direct Learning Architecture - DLA

### 2.6.2 Direct Learning Architecture

The direct learning architecture (DLA) is depicted in Fig 2.15, where the error to be minimized is directly the difference between the DPD input and the PA output normalized by a reference gain.

The DLA using Nonlinear filtered-x least mean square algorithm (NFxLMS) is proposed in [47]. The model of DPD is estimated according to the PA model and the reference error $\epsilon(n)$ which is the difference between the DPD input $u(n)$ and the normalized PA output $z(n)$. The coefficient $c_k$ of the $k$-th basis function $\Phi_k[u(n)]$ could be updated by applying the stochastic gradient algorithm, where the gradient is represented by the derivative

$$\begin{aligned}\frac{\partial \epsilon^2(n)}{\partial c_k} =& 2\epsilon^*(n)\frac{\partial \epsilon(n)}{\partial c_k} \\ =& -2\epsilon^*(n)\frac{\partial z(n)}{\partial c_k}.\end{aligned} \tag{2.40}$$

Assuming that $c_k$ vary slowly, we can have

$$\begin{aligned}\frac{\partial z(n)}{\partial c_k} =& \sum_{l=0}^{L-1} \frac{\partial z(n)}{\partial x(n-l)} \cdot \frac{\partial x(n-l)}{\partial c_k} \\ \approx& \sum_{l=0}^{L-1} \frac{\partial z(n)}{\partial x(n-l)} \cdot \Phi_k[u(n-l)]\end{aligned} \tag{2.41}$$

where $L$ is the memory depth of DPD model, and $\frac{\partial z(n)}{\partial x(n-l)}$ is the derivative of a nonlinear model of PA normalized by its gain. Thus a model of PA needs to be firstly identified.

If we denote $g(l,n) = \frac{\partial z(n)}{\partial x(n-l)}$, then (2.41) can be rewritten as

$$\frac{\partial z(n)}{\partial c_k} \approx \sum_{l=0}^{L-1} g(l,n)\Phi_k[u(n-l)]. \tag{2.42}$$

It is equivalent that each basis function $\Phi_k[u(n)]$ is filtered by an instantaneous equivalent linear (IEL) filter $\boldsymbol{g}(n) = [g(0,n),..,g(L-1,n)]$ [50]. Using this method, an NFxRLS is proposed for recursive least square (RLS) algorithm. Replacing (2.42) into (2.40), we can have

$$\begin{aligned}\frac{\partial \epsilon^2(n)}{\partial c_k} \approx & -2\epsilon^*(n)\sum_{l=0}^{L-1} g(l,n)\Phi_k[u(n-l)] \\ = & -2[\sum_{l=0}^{L-1}\epsilon(n)g^*(l,n)]^*\Phi_k[u(n-l)] \\ = & -2[\sum_{l=0}^{L-1}\epsilon(m+l-L+1)g^*(l,m+l-L+1)]^*\Phi_k[u(m-L+1)]\end{aligned} \tag{2.43}$$

if we change the variable $m = n-l+L-1$.

An adjoint IEL filter $\boldsymbol{g}_{adj}(n) = [g^*(L-1,n),..,g(0,n-L+1)^*]$ is then defined in [50]. It is equivalent that the error signal $\epsilon(n)$ is filtered by the adjoint IEL filter. By applying the adjoint IEL for LMS and RLS: Nonlinear adjoint least mean square algorithm (NALMS) and Nonlinear adjoint recursive least square algorithm (NARLS) are proposed in [50]. The computational complexity and memory requirements are reduced because only the error signal is filtered instead of every basis function of $u(n)$.

Instead of minimizing the residual between measured output and input signal, Weighted Adjacent Channel Power (WACP) is used in [51] as the objective function to minimize. WACPs represent the distortions in the lower and upper adjacent channel frequencies. It is chosen to avoid the gain/delay compensation errors and Analog Digital Convertor (ADC) distortion associated with a full time domain feedback path.

In [52], a closed-loop estimator of DLA is proposed to estimate the DPD coefficients. The advantage of this algorithm is that there is no need to identify the PA model and the identification of the coefficient errors is solved as a linear problem. Thus it is mainly explained in this section. For models linear with respect to their

coefficients, the relation between the input $\boldsymbol{u}$ and the output $\boldsymbol{x}$ of the DPD can be expressed in matrix form by

$$\boldsymbol{x} = \mathbf{U}\boldsymbol{c} \tag{2.44}$$

where $\boldsymbol{x}=[x(1),\ldots,x(N)]^T$, $\boldsymbol{c}$ is $R\times 1$ coefficient vector containing the set of $c_i$ ($i$=1,..,$R$), $\mathbf{U}$ is $N\times R$ matrix of basis functions $\Phi_R[\boldsymbol{u}]$ where $\boldsymbol{u}=[u(1),\ldots,u(N)]^T$.

The reference error of measurement is calculated by

$$\epsilon(n) = \frac{y(n)}{G} - u(n) = z(n) - u(n). \tag{2.45}$$

The origin of the reference comes from two parts:

- The coefficients of DPD are not ideal. As the DPD model is linear with its coefficients, we can express the error generated by the coefficients error $\boldsymbol{\Delta c}$ by $\mathbf{U}\cdot\boldsymbol{\Delta c}$, where $\boldsymbol{\Delta c}$ is $R\times 1$ vector containing the set of coefficient errors $\Delta c_i$.
- The LS error $\boldsymbol{\epsilon}_{LS}$ in the identification. In LS calculation, QR factorization is an orthogonal projection. Thus $\boldsymbol{\epsilon}_{LS}$ is orthogonal to the input signal $\mathbf{U}$.

Thus the error signal can be also written as

$$\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_{LS} + \mathbf{U}\cdot\boldsymbol{\Delta c}, \tag{2.46}$$

To reduce $\boldsymbol{\epsilon}_{LS} = \boldsymbol{\epsilon} - \mathbf{U}\cdot\boldsymbol{\Delta c}$, we have the cost function to minimize:

$$J = \sum_n |\epsilon(n) - \sum_{k,l} \Delta c_{k,l}\cdot\Phi_R[u(n)]|^2. \tag{2.47}$$

The LS solution of the coefficient error which minimize (2.47) is the solution for the following equation

$$\mathbf{U}^H\boldsymbol{\epsilon} = [\mathbf{U}^H\mathbf{U}]\boldsymbol{\Delta c} \tag{2.48}$$

where $\boldsymbol{\epsilon} = [\epsilon(1),\ldots,\epsilon(N)]^T$.

With the estimated coefficient error $\boldsymbol{\Delta c}^{(i)}$ at $i$-th iteration, the coefficients are updated iteratively:

$$\boldsymbol{c}^{(i+1)} = \boldsymbol{c}^{(i)} - \eta\cdot\boldsymbol{\Delta c}^{(i)} \tag{2.49}$$

where $i$ indicates the iteration number, $0<\eta\leq 1$.

In this approach, we do not need to calculate the inverse of the PA model.

## 2.7 Test bench

In order to validate the effectiveness of the proposed algorithms and criteria, experiments have been carried out using a test bench. The block diagram and the photo of the test bench are shown in Fig 2.16 and Fig 2.18 respectively.

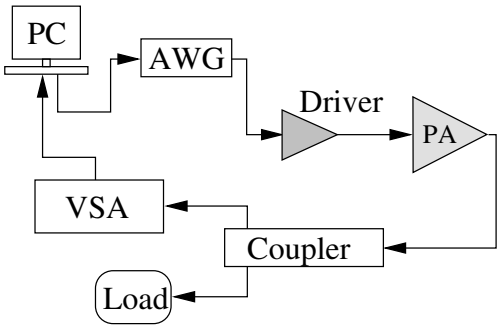
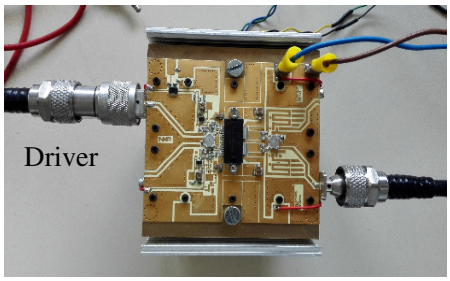


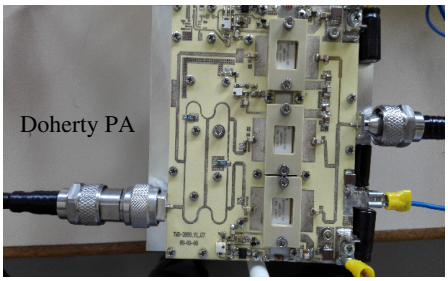

Figure 2.16: Test Bench Blocks Diagram (AWG stands for Arbitrary Waveform Generator and VSA for Vector Spectrum Analyzer)

The baseband IQ signal is fed from the PC Workstation to the PA chain through an Arbitrary Waveform Generator (AWG) using a 200 MHz sampling frequency. The AWG up-converts the baseband signal to the carrier frequency. An N5182B MXG X-Series RF Vector Signal Generator is used as AWG (carrier frequency range from 9 kHz to 6 GHz).

The signal at the output of the PA is then down-converted to baseband by a Vector Spectrum Analyzer (VSA) which provides to the PC workstation the baseband signal digitized with a maximum sampling frequency of 200 MHz. A




(a) Driver in the test bench




(b) Three-way Doherty PA

Figure 2.17: Three-way Doherty PA and Driver

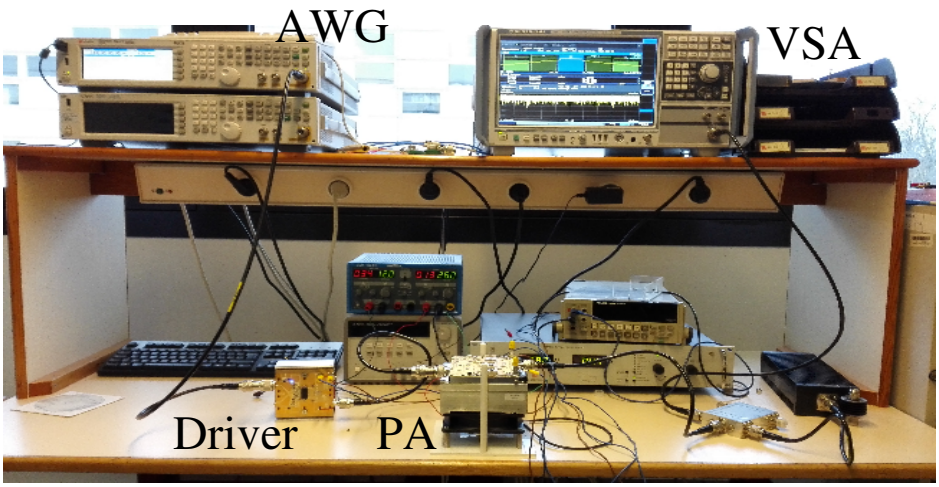



Figure 2.18: Test bench for Experimental Implementation

Rohde & Schwarz FSW signal & spectrum analyzer is used as VSA (reception frequency range between 2 Hz and 8 GHz). In this dissertation, the VSA sampling frequency used is 150 MHz.

The input and output baseband signals are then synchronized in time to be used by the identification algorithm (2.38).

Two PAs have been used to validate the proposed algorithms.

The first PA line is made of a three-way Doherty PA designed for base station (BS-PA) (Fig 2.17b) with three LDMOS transistors BLF7G22LS-130 from Ampleon, formerly NXP and its associated driver (Fig 2.17a). This Doherty PA is capable of a peak output power of 57 dBm (500 W) and has a linear gain of 16 dB. For most of the experiments in this dissertation, the stimulus signal is an LTE signal with 20 MHz bandwidth and a PAPR of approximately 8 dB, and the carrier frequency is 2.14 GHz. The modulation of the LTE signal is QPSK.

The nonlinearities and the memory effect of this PA can be seen from the AM-AM/AM-PM curves in Fig 2.19. In this example, the average power of the signal at the input of the driver is 5 dBm. The linear gain of the driver is 31.5 dB. The measured average output power of DPA is 47.3 dBm, and the measured peak power is 53.4 dBm. The spectrum of PA output captured by VSA is illustrated in Fig 2.20.

The second PA line is made of a Doherty PA designed for broadcast (BrC-PA). Its average output power is 200W. Its normalized AM/AM & AM/PM curves are depicted in Fig 3.2. The input signal is an OFDM signal with 8 MHz bandwidth

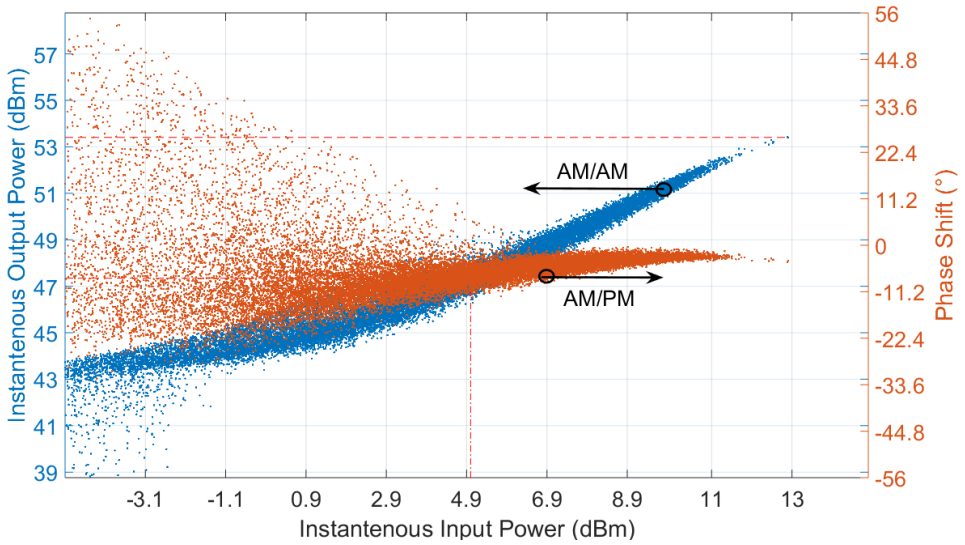



Figure 2.19: AMAM & AMPM curves of driver and Doherty PA for an LTE 20MHz input signal with 7.8 dB PAPR

and a PAPR of approximately 11 dB, and the carrier frequency is 666 MHz.

These measurements have been obtained thanks to the support of National Instruments on Digital Predistortion Framework research activity [53] and the support of Teamcast in the frame of the ambrun project (FUI AAP11) [54].

The computations described hereafter have been done on an Intel Xeon CPU E3-1245 v3 at 3.40 GHz.

## 2.8 Conclusion

This chapter introduces the generalities of digital predistortion of power amplifiers. The characteristics of PA and the motivation of applying DPD are discussed. Different models can be used as DPD.

The simplest model is a polynomial model which has a deficiency in the performance of high power amplifier predistortion with wide bandwidth signals because it cannot compensate for the memory effect. Thus more general Volterra series model is preferred in the case when memory effects of PA are not negligible. However the complexity of Volterra series model is too high to implement. Hammerstein model and Wiener model are substitutions which compensate for the

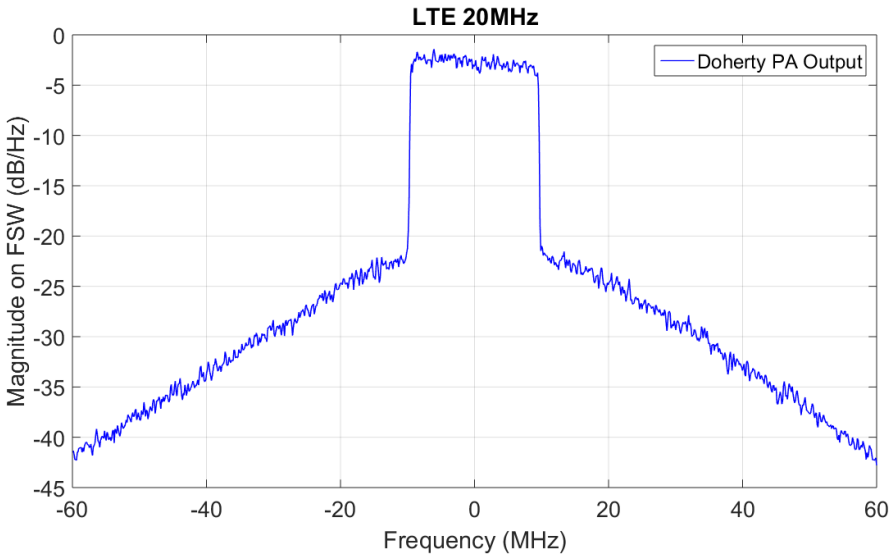



Figure 2.20: Output signal spectrum of Doherty PA excited by 20MHz bandwidth LTE signal

nonlinearity and the memory effect separately with a reasonable number of terms. For more general performance, Wiener-Hammerstein model is composed by concatenating these two models together. Memory polynomial model and generalized memory polynomial model are proposed to enhance the modeling accuracy with a reasonable increase of model complexity.

To improve DPD performance, some models are proposed by optimizing different aspects. The polynomial model can be orthogonalized to improve the performance of model coefficient estimation. Vector-switched model and Decomposed Vector Rotation model are based on vector quantification which may simplify the model when the PA characteristics is strongly nonlinear.

Multilayer perceptron neural network is different from the mathematical models above, which can also achieve a good performance in behavioral modeling.

The coefficients of the DPD model can be identified using DLA and ILA. In the end of this chapter, the test bench for the experimental implementations in this dissertation is introduced.

The main contributions of this dissertations are thoroughly presented in the following chapters.

# Chapter 3

# Determining the Structure of Digital Predistortion Models

## 3.1 Introduction

Volterra series has good performances in PA or DPD modeling. However its complexity is very high. Some models are derived from Volterra series by applying different pruning techniques. The number of basis functions can be reduced by removing those terms which have very few influences on modeling accuracy. Thus an optimal structure of the model which has low complexity while keeping also a good linearization performance is needed.

There are different methods which optimize the structure to have a good trade-off between modeling accuracy and model complexity. Some basis function selection techniques to make the model sparse are discussed in this chapter. Another way to prune the model is to cut off the basis functions with high orders and deep memories. Thus some optimization algorithms can be also applied on this problematic to optimize the model dimension. The state of the art is presented in the following sections.

In this chapter we present the first contribution of this dissertation which is an algorithm based on hill-climbing heuristics to determine an optimal model structure of DPD according to some criteria. These criteria represent trade-offs between modeling accuracy and model complexity. The performance of this algorithm is verified on test bench and is compared with other optimization algorithms. The advantage of the proposed algorithm is that its search path can be controlled and some optimizations can then be made to reduce the execution time. The GMP

model is taken as an example in the implementations.

This chapter is organized as follows:

- Section 2 presents a bibliographic study in model structure optimization.
- The proposed search algorithm based on Hill-Climbing is described in Section 3.
- In Section 4, we propose two different search criteria for Hill-Climbing heuristic.
- Section 5 presents different pruned neighborhoods to reduce the execution time.
- In Section 6, two methods to estimate the weighting coefficient of the criteria are explained.
- The experimental results are presented and discussed in Section 7.
- Section 8 compares the performances of hill-climbing and genetic algorithms.
- Section 9 introduces First-choice hill-climbing which can reduce the execution time while keeping the same performance.
- A brief conclusion is given in Section 10.

## 3.2 Bibliographic study

Behavioral modeling, which is known as black-box modeling, is a very high level modeling as it is based only on the observation of circuit input and output signals. Even with *a priori* knowledge of the PA internal composition, it is difficult to determine the structure of a behavioral model which has low complexity and high performance. There are mainly two methods: selection of basis functions and optimization algorithms for model structure determination.

### 3.2.1 Basis Functions Selecting

The complexity of model identification using (2.38) and (2.48) depends on the number of basis function $R$. Some pruning technique to reduce the number of basis function while keeping the same modeling performance have been studied [55] [56] [57].

An adaptive scheme for selecting basis functions in stochastic conjugate gradient (SCG) computations is proposed in [55].

In matrix calculation of SCG, the basis functions are selected according to their corresponding residual values $\boldsymbol{r} = [\gamma_1, \cdots, \gamma_R]$. The residual $\gamma_i$ ($i$=1,..,$R$) of each basis function $f_i$ is defined as the scalar product $\frac{1}{N}(\boldsymbol{f}_i \cdot \boldsymbol{\epsilon})$ where $\boldsymbol{f}_i = [f_i(1), \cdots, f_i(N)]$, $\boldsymbol{\epsilon}$ is the $N$-sample error vector. For ILA, the error is the difference between the predistorted signal $x(n)$ and estimated signal $z_p(n)$ in Fig 2.14; and for DLA, the error is the difference between the original signal $u(n)$ and the feedback signal $z(n)$ when DPD is present as in Fig 2.15 [56].

When the residual is negligible, it means the corresponding basis function is not important and can be omitted. The actual value of residual is registered as the residual value of that basis function. The average value of the residuals of the set of basis functions are set as a threshold. The basis function with residual higher than the threshold is selected. The SCG uses different sample sets of the input signal $u(n)$ at different iterations. Thus the threshold is adjusted periodically according to the new residual calculated in new iterations because the importance of basis functions varies for different sample sets. In the preconditioned stochastic gradient method (PSGM) proposed in [58] based on SCG, a constraint is added to limit the change in derivatives from one sample set to another.

The advantage of this adaptive scheme with SCG is that the complexity is reduced and, at the same time, the convergence rate and the stability of iteration are increased. However this method to calculate the residual of basis functions is not very stable because the different basis functions present collinearity. Error Variation Ranking (EVR) method has been proposed to alleviate the effect of collinearity of basis functions [59].

In EVR method, the importance of a basis function is evaluated by the difference between the modeling performances of DPD model with this basis and without this basis. The modeling performance is estimated by the normalized mean square error (NMSE) value (2.7) between the predicted and measured output signals. The basis functions are ranked according to their importances and only some of the most important basis function will be taken. The NMSE variation caused by removing a basis function is used as quantification factor of its

importance. This helps overcoming the problem that potential multicollinearity between basis functions of MP model influences the pruning effects on prediction precision.

In [60], a technique to prune the MP model structure is proposed. The aim is to minimize the number of kernels while the residual between measured signal and predicted signal is kept in a tolerable range. The total number of kernels is decided by the nonlinearity order and the memory depth. Assuming that the kernel solution is sparse, we can keep only few of the coefficients active. The inactive coefficients are set to zero. The active coefficients of the model is estimated with a maximum likelihood method. With different groups of active coefficients, the models performances are evaluated. A technique combining Orthogonal Matching Pursuit (OMP) method and Bayesian Information Criterion (BIC) is used as criterion to evaluate their performances. The best model is then determined according to this criterion.

For a nonparametric identification model, a pruning technique is proposed in [61]. The model structure is not decided *a priori*. In this case, the characteristic curve of PA is cut into several intervals according to the input amplitude. A static nonlinear function is used as the model kernel to describe the curve by averaging the amplitude of output samples in each interval. The input signal $u(n)$ is orthogonalized to remove the correlation between $u(n)$ and $u(n-l)$ using the Gram-Schmidt (GS) process. With the orthogonalized input matrix, a static nonlinear function of each kernel which represents its importance can be calculated. The noncontributing ones are eliminated from the model structure and the modeling performance can be retained.

### 3.2.2 Model Structure Optimization Algorithms

The model structure with good trade-off between modeling accuracy and model complexity can be also determined by optimization algorithms. There have been few studies on sizing nonlinear models in DPD implementation [62] [63] [64] [65].

If we take the example of GMP model as (2.16), there are 8 sizing parameters: the nonlinearity orders ($\mathcal{K}_a$, $\mathcal{K}_b$ and $\mathcal{K}_c$), and memory depths ($\mathcal{L}_a$, $\mathcal{L}_b$, and $\mathcal{L}_c$), and the lagging and leading delay tap lengths ($\mathcal{M}_b$ and $\mathcal{M}_c$). As these parameters can have their values changed independently, they may compose an 8-dimension discrete space of GMP model structures. An exhaustive search could be used to size the GMP model but it may be too time consuming. An inadequate number of basis functions can result in insufficient accuracy or over-fitting problems. The

difficulty of sizing GMP model is due to the very large number of different possible model structures. The model basis functions have impacts on each other because of nonorthogonality. New identification of model coefficients is needed even if only one basis function is added or removed. Adding or removing some basis functions may have a predictable effect on model complexity, but an uncertain influence on modeling accuracy.

For illustration purposes, let us consider the sizing of an MP model which has only two sizing parameters: the order of nonlinearity $\mathcal{K}$, and the memory depth $\mathcal{L}$. The basis functions in this case are in the form $x(n-l)|x(n-l)|^k$, where $k \in [0, \mathcal{K}-1]$ and $l \in [0, \mathcal{L}-1]$. If $\mathcal{K} = 4$ and $\mathcal{L} = 3$ and the terms of $(k = 0, l = 0)$ and $(k = \mathcal{K}-1, l = \mathcal{L}-1)$ always exist, we have the following possible arrays of nonlinear orders: $\{0, 1, 2, 3\}$, $\{0, 1, 3\}$, $\{0, 2, 3\}$, $\{0, 3\}$, and memory taps: $\{0, 1, 2\}$, $\{0, 2\}$. In this case, both full and sparse arrays are considered since it may not be necessary to implement all nonlinear orders or delay taps over the given ranges. Thus there are a total of 8 possible model structures in this case. In general, for each combination of $\mathcal{K}$ and $\mathcal{L}$, there are $2^{\mathcal{K}+\mathcal{L}-4}$ possible model structures. By increasing the values of $\mathcal{K}$ and $\mathcal{L}$, the number of possible solutions will increase geometrically.

In this chapter we will consider only the discrete space of GMP structures with full arrays. For the GMP model which has 8 sizing parameters, an exhaustive search is very time consuming. For instance if all parameters of the GMP model are bounded between 1 and 10, there are $10^8$ models to test using exhaustive search when only full arrays are considered, which is very computationally demanding and time consuming.

A general sweep method, in which only full arrays have been considered, has been used to size MP model in [66] and [67]. All possible combinations of $\mathcal{K}$ and $\mathcal{L}$ have been evaluated in order to assess the models with different trade-offs between modeling accuracy and complexity. The one which conforms user demands the most is chosen as the best model. However other models with more sizing parameters, e.g. GMP, cannot be sized with this method. To date, their parameters are most of the time determined empirically by adjusting one or more parameter values until the adequate performance is achieved [21] [38] [68]. And there is no guarantee that the model found in this way achieves the best trade-off between modeling accuracy and complexity.

A search algorithm is needed to find the optimal GMP model structure [62]. Optimization algorithms have been widely used in analog electronic domain. Genetic Algorithms (GA) have been applied for microwave circuit models [69]. In digital predistortion, it has been used to identify the coefficients of DPD [70] [71].

In the sizing of analog circuits, not only GA [72] [73] but also Particle Swarm Optimization (PSO) [74] have already been shown to be effective. There are also a large number of parameters, constraints and performances to handle. Though some studies on optimizing DPD architectures with PSO [63] [64] or with evolutionary algorithms in [75] have been done, the DPD models used have much less parameters to determine than GMP model.

PSO is used to estimate the dimension of MP model DPD in [63]. A system is proposed to achieve the best trade-off between accuracy and number of zeros in the coefficients' vector. PSO simulates the social behavior of a particle swarm, e.g. swarm of birds or fish. The particles approach the optimum by iteratively improving the candidate solutions according to their positions and velocities.

A method using an integer GA to find the 8 sizing parameters of the GMP model has been proposed in [62]. In this algorithm, a fitness function which represents the trade-off between modeling accuracy and structure complexity is used as a search criterion. Compared with the exhaustive search, a solution close to the optimal model structure can be found much faster.

Simulated annealing (SA) is another algorithm for optimization problems, which simulates the process of annealing in metallurgy. This method is a metaheuristic which accepts not only the best solutions in order to overcome the local optima. The global optimum will be approached if a very long execution time is allowed. However, the execution time is critical on sizing the model structure, especially when the procedure is implemented in real time.

## 3.3 Hill-Climbing Heuristic

Though PSO and GA are good methods of optimization, it is difficult to control and interpret the path followed by the search algorithm because of its random nature. Compared with PSO and GA, the advantage of HC is that the search path can be controlled by configuring the neighborhood definition. To the best of the author knowledge, this is the first time that an algorithm based on Hill-Climbing for sizing a GMP model DPD structures has been proposed and studied.

Hill-Climbing algorithm [76] [77] is a heuristic which is an iterative algorithm that begins from an initial solution, then attempts to find a better solution by comparing the current solution with its neighbors. If there is a better solution among the neighbors, it is taken as the new solution. The search procedure repeats until no better solution can be found.

Instead of exhaustive search, we apply an iterative search (Hill-Climbing al-

gorithm [76]) and test only a small number of models in each iteration. The tested models in each iteration can be selected in different ways. In this chapter, we select the neighbors of the solution of each iteration to test. The neighbors have also different definitions because the search space is multi-dimensioned. The algorithm depends on two factors: the expression of merit function and the definition of neighborhood used for search.

In the discrete space $U$ which embodies all different GMP model structures, each element $x_i$ can be assigned to a unique structure. The element $x_i$ consists of 8 integer coordinates: $\mathcal{K}_{a,i}$, $\mathcal{L}_{a,i}$, $\mathcal{K}_{b,i}$, $\mathcal{L}_{b,i}$, $\mathcal{M}_{b,i}$, $\mathcal{K}_{c,i}$, $\mathcal{L}_{c,i}$, $\mathcal{M}_{c,i}$, which are the sizing parameters of the model structure. The value of a merit function $J(x_i)$ is associated to each element $x_i$.

Hill-Climbing heuristic tests only the elements which are the neighbors of the solution at each iteration. In our implementation, we define a neighbor of element $x_i$ as an 8-tuple ($\mathcal{K}_{a,i} + \delta_1$, $\mathcal{L}_{a,i} + \delta_2$, $\mathcal{K}_{b,i} + \delta_3$, $\mathcal{L}_{b,i} + \delta_4$, $\mathcal{M}_{b,i} + \delta_5$, $\mathcal{K}_{c,i} + \delta_6$, $\mathcal{L}_{c,i} + \delta_7$, $\mathcal{M}_{c,i} + \delta_8$), where $\delta_{1,\dots,8} \in [0, \pm 1]$ and $\delta_{1,\dots,8}$ are not 0 at the same time. The subspace of neighbors is denoted by $S$, $S \subset U$.

**Algorithm 1:** Overview of Hill Climbing Heuristic

Set the loop counter $q$=1;
Choose an 8-tuple element $x_{0(1)}$ as the initial element;
Evaluate the initial element merit value $J(x_{0(1)})$;
$x_{s(0)} = x_{0(1)}$;
**while** *(1)* **do**
  Compute all $J(x_{i(q)})$ where $x_{i(q)} \in S_{(q)}$;
  $x_{s(q)} = \arg\min_{x_{i(q)} \in S_{(q)}} \left(J(x_{i(q)})\right)$;
  **if** $J(x_{s(q)}) < J(x_{s(q-1)})$ **then**
    $q = q + 1$;
    $x_{0(q)} = x_{s(q-1)}$;
  **else**
    end while loop;
  **end**
**end**
Take the last solution $x_{s(q)}$ as the best element;

As Algorithm 1 shows, the Hill-Climbing heuristic starts from a given $x_{0(1)}$ at the first iteration and continually moves in the direction of the element with the best merit value. At the $q$-th iteration, the search procedure starts from $x_{0(q)}$ and

test the subspace of its neighbors $S_{(q)}$. The neighbors are denoted by $x_{i(q)}$, where $x_{i(q)} \in S_{(q)}$. The merit value $J(x_{i(q)})$ associated to $x_{i(q)}$ is evaluated. The solution $x_{s(q)}$ is the element with the minimum merit value among the neighbors.

To simplify the notations, we denote the merit value $J(x_{i(q)})$ by $J_{i(q)}$.

We compare the merit value of the solution $x_{s(q)}$ with the solution of the previous iteration $x_{s(q-1)}$. If $J_{s(q)}$ is lower than $J_{s(q-1)}$, at the following iteration, the search procedure starts from $x_{s(q)}$ and tests its neighbors. Otherwise the algorithm stops and $x_{s(q)}$ is taken as the final solution. As the neighbors of $x_{0(q)}$ and $x_{0(q+1)}$ are partially overlapped, *i.e.* $S_{(q)} \cap S_{(q+1)} \neq \emptyset$, we do not evaluate the elements which have been already evaluated at the previous iterations. The search ends when the solution is not changed in the latest iteration.

## 3.4 Search Criteria

Improving modeling accuracy by increasing the number of basis functions induces inevitably an increase in model complexity [56]. Thus the trade-off between modeling accuracy and model complexity is a very important criterion for DPD model selection.

The merit value in our algorithm refers to a certain criterion which leads to a trade-off between modeling accuracy and model complexity. $\text{NMSE}_{\text{dB}}$ (2.7) can be used to evaluate modeling accuracy. It will be denoted by $Y$ hereafter. The model complexity is represented by the number of coefficients of the model, denoted by $R$.

We can also take the complexity of model identification instead of model complexity as a criterion. The complexity of model identification is in order of $R^2$ if we consider solving (2.38) using a QR decomposition as described in Section 2.6.1.

In the following sections, $\text{NMSE}_{\text{dB}}$, number of coefficients and the merit value of element $x_i$ are denoted by $Y_i$, $R_i$ and $J_i$ respectively.

### 3.4.1 Weighted Combination of Objectives

The aim of combining two objectives into one criterion is for example to find the model with minimum number of coefficients $R$ while its modeling accuracy $Y$ is kept in a tolerable range.

In [62], a fitness function is defined as a weighted additive combination of accuracy and the complexity which is represented by the number of coefficients.

In our work, we have considered two different ways to combine $Y$ and $R$ into a single hybrid criterion: additive and multiplicative combinations.

For the additive combination, the resulting criterion has the same form as in [62] [65]

$$J_i = Y_i + \mu R_i, \tag{3.1}$$

where $\mu$ is a positive real value.

For the multiplicative combination

$$J_i = Y_i \left(1 - \alpha \cdot R_i\right) \tag{3.2}$$

which can be rewritten in

$$J_i = Y_i + \beta(Y_i) R_i \tag{3.3}$$

where $\beta(Y_i) = -\alpha \cdot Y_i$ is a dynamic weighting coefficient compared to $\mu$ in (3.1). The value of $\alpha$ should respect $(1 - \alpha \cdot R_i) > 0$. Thus $\frac{1}{\alpha}$ is the maximum of number of coefficients.

If we take the complexity of model identification for complexity evaluation, the criteria are kept the same except $R_i$ is replaced by $R_i^2$ in (3.1) and (3.3).

### 3.4.2 Additive Criterion

While comparing two elements $x_i$ and $x_j$ in the whole space, we denote the differences of $\text{NMSE}_{\text{dB}}$ and number of coefficients between $x_i$ and $x_j$ by $\Delta Y_{ij}$ and $\Delta R_{ij}$ respectively:

$$\begin{aligned} \Delta Y_{ij} &= Y_i - Y_j \\ \Delta R_{ij} &= R_i - R_j, \end{aligned} \tag{3.4}$$

For the additive criterion, two elements $x_i$ and $x_j$ are considered as equivalent if:

$$\begin{aligned} Y_i + \mu R_i &= Y_j + \mu R_j \\ \text{or equivalently } Y_i - Y_j &= -\mu (R_i - R_j) \end{aligned} \tag{3.5}$$

According to (3.4) and (3.5), we obtain

$$\begin{aligned} \Delta Y_{ij} &= -\mu \Delta R_{ij} \\ \text{or } \frac{\Delta Y_{ij}}{\Delta R_{ij}} &= -\mu. \end{aligned} \tag{3.6}$$

We make the trade-off with the additive criterion that an increment of the number of coefficients by an amount $\Delta R_{ij}$ with $R_i > R_j$ should bring the corresponding improvement of $\text{NMSE}_{\text{dB}}$ $|\Delta Y_{ij}|$ equal to or better than $\mu \Delta R_{ij}$, i.e. $|\Delta Y_{ij}| \geqslant \mu \Delta R_{ij}$. Equivalently, we can say that we accept to degrade the $\text{NMSE}_{\text{dB}}$ by an amount $\Delta Y_{ij} = Y_i - Y_j > 0$ if the reduction in number of coefficients $|\Delta R_{ij}|$ is equal to or better than $\frac{\Delta Y_{ij}}{\mu}$, i.e. $|\Delta R_{ij}| \geqslant \frac{\Delta Y_{ij}}{\mu}$.

From (3.6), $\mu$ is a constant which represents the $\text{NMSE}_{\text{dB}}$ tolerance per coefficient, or unit tolerance.

### 3.4.3 Multiplicative Criterion

Two elements $x_i$ and $x_j$ are considered as equivalent if:

$$Y_i(1 - \alpha R_i) = Y_j(1 - \alpha R_j) \tag{3.7}$$

According to (3.4) and (3.7), we have

$$\begin{aligned} \Delta Y_{ij} =& \alpha \left( Y_i R_i - (Y_i - \Delta Y_{ij})(R_i - \Delta R_{ij}) \right) \\ =& \alpha (Y_i \Delta R_{ij} + \Delta Y_{ij} R_i - \Delta Y_{ij} \Delta R_{ij}) \\ =& \alpha Y_i \Delta R_{ij} + \alpha \Delta Y_{ij} (R_i - \Delta R_{ij}). \end{aligned} \tag{3.8}$$

The normalized variation of $\text{NMSE}_{\text{dB}}$ is formulated as:

$$\frac{\Delta Y_{ij}}{Y_i} = \frac{\Delta R_{ij}}{\frac{1}{\alpha} - (R_i - \Delta R_{ij})}. \tag{3.9}$$

In practical case as we will see in the Section 3.7, the value of $\frac{1}{\alpha}$ is much greater than $(R_i - \Delta R_{ij})$. We can neglect the term $(R_i - \Delta R_{ij})$ and make the following approximation of (3.9) as:

$$\frac{\Delta Y_{ij}}{Y_i} \approx \alpha \Delta R_{ij}. \tag{3.10}$$

When $|Y_i|$ gets large, the acceptable value of $|\Delta R_{ij}|$ decreases. In other words, when the modeling is more accurate, we have more requirements on number of coefficients.

For example, if $\alpha = 0.002$, to have an improvement of $\text{NMSE}_{\text{dB}}$ $|\Delta Y_{ij}|$=1 dB when $\text{NMSE}_{\text{dB}}$ is around -10 dB, we accept a GMP model with 50 coefficients more. When $\text{NMSE}_{\text{dB}}$ is around -50 dB, we accept a GMP model with only 10 coefficients more. The main idea of this criterion is that when the modeling accuracy is low, we take it as the major factor, and when it is satisfactorily high, we take the model complexity as the major factor.

## 3.5 Weighting Coefficient Determination

In this chapter, we propose two methods to determine the weighting coefficient $\mu$ or $\alpha$: off-line computation and on-line computation, using a sub-space of the discrete space $U$ of GMP model structures.

As described in previous subsections, $\mu$ could be determined by giving certain values of $\frac{\Delta Y_{ij}}{\Delta R_{ij}}$ in (3.6), and $\alpha$ could be determined by giving directly a certain percentage value depending on the constraint of the application.

And the approach is to consider the global optimum that can be defined in the following objective as shown in Fig 3.1:

$$\begin{aligned} &\min R, \text{ then } \min Y \\ &\text{subject to } Y \leqslant Y_{th}. \end{aligned} \tag{3.11}$$

where the threshold $Y_{th}$ is the worst $\text{NMSE}_{\text{dB}}$ value that we tolerate and its value depends on the lowest $\text{NMSE}_{\text{dB}}$ that a GMP model can reach. If there are more than one element with the same number of coefficients satisfying the condition, the one with the lowest $\text{NMSE}_{\text{dB}}$ value is selected. However, it is impossible to know the lowest $\text{NMSE}_{\text{dB}}$ for GMP without testing all possible models. Thus the criteria (3.1) and (3.2) are used to replace (3.11) with correct values of $\mu$ or $\alpha$ determined by the proposed methods.

If we denote this global optimum by $x_p$, $\mu$ or $\alpha$ is determined in order that the merit value of $x_p$ is better than any other element. Thus, for all elements $x_n$, $n \neq p$. We have $J_n \geqslant J_p$, which can be written as:

$$\mu(R_n - R_p) \geqslant -(Y_n - Y_p). \tag{3.12}$$

for the additive criterion, and

$$\alpha(Y_n R_n - Y_p R_p) \leqslant Y_n - Y_p. \tag{3.13}$$

for the multiplicative criterion.

For a certain value of $Y_{th}$, there is a corresponding $\mu$ or $\alpha$ to make the combined criterion (3.1) or (3.2) realize (3.11). Off-line and on-line computation are used to estimate their values by sampling the whole space.

At the $q$-th iteration, $\text{NMSE}_{\text{dB}}$ and number of coefficients of element $x_{i(q)}$ are denoted by $Y_{i(q)}$ and $R_{i(q)}$ respectively.

As the determinations of $\mu$ in (3.1) and $\alpha$ in (3.2) are identical, we explain only determination of $\mu$ in following subsections.

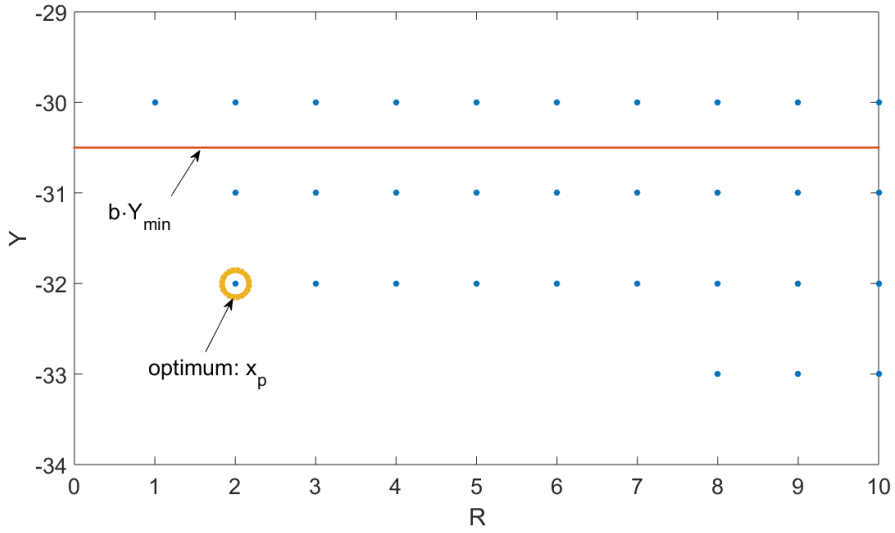



Figure 3.1: Global Optimum

### 3.5.1 Off-line Computation

In this method, $\mu$ is computed off-line before the search with respect to (3.11).

We choose a subset that includes all MP model structures with memory order $\mathcal{K}_a$ bounded by $\mathcal{K}_a \leqslant \mathcal{K}_{a,max}$ and memory length $\mathcal{L}_a$ bounded by $\mathcal{L}_a \leqslant \mathcal{L}_{a,max}$. This subset contains $\mathcal{K}_{a,max}\mathcal{L}_{a,max}$ MP structures (which is much smaller than the number of GMP structures).

We calculate the $\text{NMSE}_{\text{dB}}$ for all the elements in this subset. We fix the threshold of $\text{NMSE}_{\text{dB}}$, $Y_{th}$, equal to $bY_{min}$ with $bY_{min} > Y_{min} = b \cdot Y_{min}$, where $Y_{min}$ is the minimum $\text{NMSE}_{\text{dB}}$ value obtained for MP models and $0 < b < 1$ is a constant. For a given $b$, we get the element $x_p$ which has the minimum number of coefficients $R_p$ and $Y_p \leqslant Y_{th}$ The element $x_p$, with $p \in [0, \mathcal{K}_{a,max}\mathcal{L}_{a,max} - 1]$, is regarded as the best element. $\mu$ is determined in order that the merit value of $x_p$ is better than any other element. Thus, several inequalities are constructed for all elements $x_n$, with $n \neq p$. We have $J_n \geqslant J_p$, which gives (3.12).

The estimation of $\mu$ is described in Algorithm 2. As $(R_n - R_p)$ can be positive or negative, we can get an interval of $\mu$. The middle value of the interval will be then taken as $\mu$ for the search in the full space of GMP structures.

### 3.5.2 On-line Computation

In this method, $\mu$ is computed on-line during the search. The chosen subspaces are composed of the GMP models tested at each iteration (at $q$-th iteration, the subspace is $S_{(q)}$). The number of models in the subset at $q$-th iteration is $F_{(q)}$.

At the first iteration, $\mu$ is not estimated and the merit function does not exist. The solution $x_{s(1)}$ is decided by the following steps:

- find $Y_{min(1)}$
- select all elements with $Y \leqslant b \cdot Y_{min(1)}$
- find the element with the minimal $R$ among these elements
- then find the element with the minimal $Y$ if multiple elements found in last step

where $Y_{min(1)}$ is the lowest $Y$ at the first iteration.

To calculate $\mu$, at each iteration, we select a local optimum $x_{p(q)}$ exactly with the same method as the method described in off-line computation. The only difference is that the subspace here is the neighborhood tested at each iteration instead of MP models.

Considering $x_{p(q)}$ as the element with the highest merit value in the $q$-th iteration, $F_{(q)}$ inequalities as (3.12) are constructed for each element $x_{n(q)}$, with $n \in [0, F_{(q)}], n \neq p$. Thus at each iteration, we can have an interval of $\mu$ and we take the middle value.

On-line computation allows to benefit from a larger sample size than off-line computation, the actual $\mu$ is updated using a damping factor over the previous values by.

$$\mu^{(q+1)} = \mu^{(q)} + \gamma \times (\mu_{computed} - \mu^{(q)}). \tag{3.14}$$

The value of $\gamma$ controls the convergence trend of estimation of $\mu$. When $\gamma$ is too large, the value of $\mu$ fluctuates much. When $\gamma$ is too little, the convergence speed is very slow. In the experiments, a damping coefficient $\gamma = 0.6$ is used.

## 3.6 Pruned Neighborhoods

Since we noticed that the complexity of the algorithm depends on the number of tested elements, several methods are proposed to reduce the complexity by pruning the neighborhoods.

**Algorithm 2:** Estimation of $\mu$ interval

Initialize upper bound $ceiling = \inf$;
Initialize lower bound $floor = 0$;
**while** $n \in S, R_n \neq R_p$ **do**
    $\mu(R_n - R_p) \geqslant -(Y_n - Y_p)$;
    $\mu_{limit} = \frac{-(Y_n - Y_p)}{R_n - R_p}$;
    **if** $\mu_{limit} \in [floor, ceiling]$ **then**
        **if** $R_n > R_p$ **then**
            $floor = \mu_{limit}$;
        **else**
            $ceiling = \mu_{limit}$;
        **end**
    **end**
**end**

Literally, the neighborhood refers to the individuals with addresses in the same region or nearby. By applying different constraints on neighborhood, the size of neighborhood can be controlled. To reduce the number of neighbors, we proposed several pruned neighborhoods in this section. The constraints can be directly on the 8 sizing parameters, or on the total number of coefficients.

### 3.6.1 Constraint on Number of Coefficients

The numbers of coefficients in the neighbor may have huge variations. We propose a variation limit $d$ for the number of coefficients. In the neighborhood $S_{(q)}$ of the element $x_{0(q)}$ at $q$-th iteration, we test only those elements $x_{i(q)}$ with $|R_{i(q)} - R_{0(q)}| \leqslant d$. The set of this constrained neighborhood is denoted by $C_{(q)}$ and $C_{(q)} \subset S_{(q)}$. In this way, we can reduce the number of tests in each iteration. When there is no neighbor element better than $x_{0(q)}$, the search is paused.

In order to keep the algorithm performance, we increment the value of the variation limit $d$ once the search is paused. The neighborhood is broadened to make sure that a better element is not missed because of the pruning technique. In terms of complexity, we allow only one increment of $d$ for the same element. If no better element is found in the broadened neighborhood after the increment of $d$, we take this element as the best solution and the search terminates.

### 3.6.2 Jumping on Number of Coefficients

Considering that $C_{(q)}$ still includes many elements, we propose a jumping distance $s$ for number of coefficients. In the neighborhood $S_{(q)}$ of the element $x_{0(q)}$ at $q$-th iteration, we test only those elements $x_{i(q)}$ with $|R_{i(q)} - R_{0(q)}| = s$. The set of this constrained neighborhood is denoted by $D_{(q)}$ and $D_{(q)} \subset C_{(q)}$.

The algorithm will move forward by jumping over a large amount of elements, which makes it converge to the global optimum very fast.

With the same idea, once the algorithm is stuck at an element, we increment the value of the jumping distance $s$ to detect if there are better elements in the broadened neighborhood. This increment is applied only one time for one solution. When the search terminates, we take the current solution as the final solution.

### 3.6.3 Unidimensional neighbor

The pruning technique can be also applied on the sizing parameters. In the conventional neighborhood definition, all the parameters of a neighbor of element $x_i$ can have a variation of $\pm 1$. Thus in case of $8$ parameters, the maximum size of neighborhood is $3^8 - 1$.

To reduce the size of neighborhood, we propose to limit the the number of varied parameters to one. The neighbors of $x_i$ have parameters $\mathcal{K}_{a,i} + \delta_1$, $\mathcal{L}_{a,i} + \delta_2$, $\mathcal{K}_{b,i} + \delta_3$, $\mathcal{L}_{b,i} + \delta_4$, $\mathcal{M}_{b,i} + \delta_5$, $\mathcal{K}_{c,i} + \delta_6$, $\mathcal{L}_{c,i} + \delta_7$, $\mathcal{M}_{c,i} + \delta_8$, where $\sum_{i=1}^{8} |\delta_i| = 1$. The number of neighbors of an element decreases sharply to $16$.

This definition of neighborhood is called unidimensional neighborhood while the conventional one without constraint is called full-dimensional neighborhood. This helps the search path to converge to the region of the global optimum with very few tested models.

Once the search path is stuck at an element, we will remove this constraint and test its full-dimensional neighbors. If no better element is found, we take the current solution as the final solution.

## 3.7 Experiments and Results

### 3.7.1 Experimental Signal Acquisition

As indicated in Chapter 2, two PAs have been used to validate the proposed algorithm. The first PA line is made of the Doherty PA (Fig 2.17b) introduced in Chapter 2, which is designed as a base station PA (BS-PA).

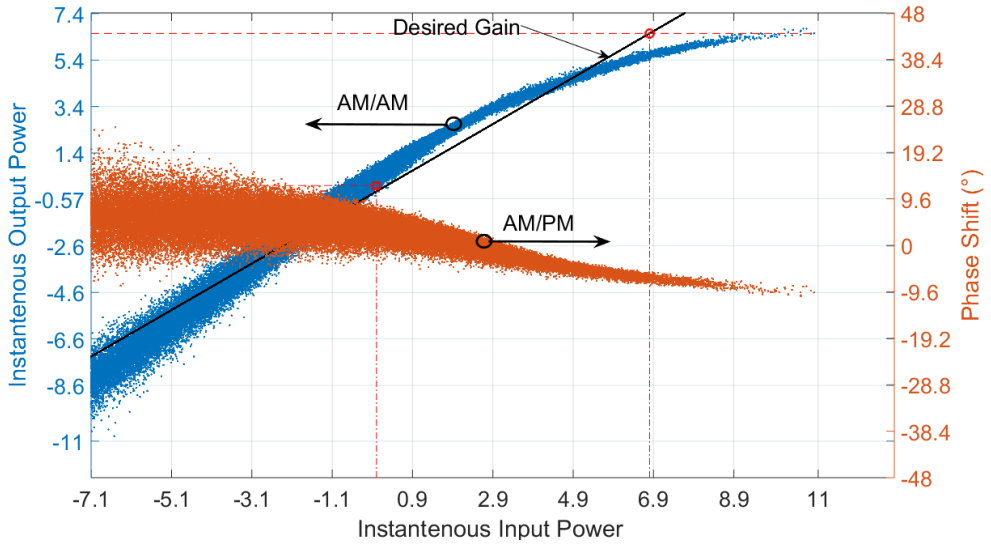



Figure 3.2: AMAM & AMPM curves of Doherty Broadcast PA

The second PA line is made of a Doherty Broadcast PA. The measured data of this PA were provided by Teamcast group. Its normalized AM/AM & AM/PM curves are depicted in Fig 3.2.

The input and output baseband signals are synchronized in time to be used by the identification algorithm (2.38).

The following subsections present the experimental results.

For each PA, base station (BS-PA) and broadcast (BrC-PA), the results searched with the following criteria are presented:

1. the additive criterion with $\mu$ determined off-line (ACoff)
2. the additive criterion with $\mu$ determined on-line (ACon)
3. the multiplicative criterion with $\alpha$ determined off-line (MCoff)
4. the multiplicative criterion with $\alpha$ determined on-line (MCon)

In each result, there are two different implementations of the search algorithm: the first implementation (1-step) is to search the best GMP model from the very beginning; and the second implementation (2-step) is divided into two steps: first we find out the MP model minimizing the search criterion; then we start from this MP model, to find the best GMP model.

The 1-step search starts from the element which corresponds to the linear function $y_{GMP}(n) = ax(n)$ where

$$\begin{aligned} \mathcal{K}_a =& 1, \mathcal{L}_a = 1 \\ \mathcal{K}_b =& 0, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 0, \mathcal{L}_c = 1, \mathcal{M}_c = 1. \end{aligned} \tag{3.15}$$

In our searches, we estimate $\mu$ and $\alpha$ with $b = 97\%$ in (3.11). We restrict the parameters to the following intervals:

$$\begin{aligned} & 1 \leqslant \mathcal{K}_a \leqslant 11, 1 \leqslant \mathcal{L}_a \leqslant 5 \\ & 0 \leqslant \mathcal{K}_b \leqslant 10, 1 \leqslant \mathcal{L}_b \leqslant 5, 1 \leqslant \mathcal{M}_b \leqslant 4 \\ & 0 \leqslant \mathcal{K}_c \leqslant 10, 1 \leqslant \mathcal{L}_c \leqslant 5, 1 \leqslant \mathcal{M}_c \leqslant 4, \end{aligned} \tag{3.16}$$

with a global constraint on the total number of coefficients $R < 80$. Each GMP model is identified using 15000 samples.

In off-line computation, the interval of $\mu$ or $\alpha$ is determined as described in Section 3.5, with $\mathcal{K}_{a,max} = 11$ and $\mathcal{L}_{a,max} = 11$.

In on-line computation, the interval of $\mu$ or $\alpha$ is computed with the tested elements at every iteration. The middle of the interval is taken to renew the value of $\mu$ or $\alpha$.

### 3.7.2 Exhaustive Search

In order to evaluate our proposed criteria (3.1) and (3.2), we have conducted an exhaustive search as a reference to get all $(Y, R)$ couples. Inspection of these results allows to determine the best GMP model (i.e. minimum number of coefficients) for a given $\text{NMSE}_{\text{dB}}$ requirement.

For the BS-PA, the lowest $\text{NMSE}_{\text{dB}}$ is -33.91 and is obtained with 79 coefficients. As shown in Fig 3.3, we have the threshold $bY_{min}$=-32.90 dB in (3.11) with $b = 0.97$. From that we get $x_p = (Y, R) = (-32.96, 43)$ with structure

$$\begin{aligned} \mathcal{K}_a =& 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 3, \mathcal{M}_b = 3 \\ \mathcal{K}_c =& 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1. \end{aligned} \tag{3.17}$$

For the BrC-PA, the lowest $\text{NMSE}_{\text{dB}}$ is -41.79 also with 79 coefficients. We have the threshold $bY_{min}$=-40.54 dB. From that we get $x_p = (Y, R) = (-40.56, 28)$

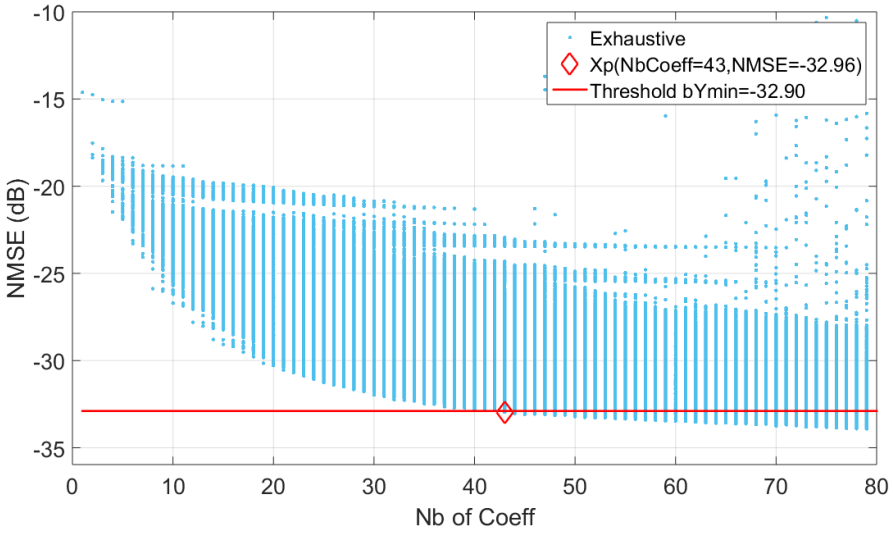



Figure 3.3: Determination of upper bound and lower bound of $\mu$ with a concave edge

with structure

$$\begin{aligned} \mathcal{K}_a =& 9, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 3, \mathcal{L}_c = 3, \mathcal{M}_c = 1. \end{aligned} \tag{3.18}$$

In the following paragraph we will use the results of the exhaustive search to compute reference values for $\mu$ and $\alpha$ as described in Section 3.5.1. Considering $x_p$ as the element with the best merit value, we can construct inequalities as (3.12) or (3.13) to determine an interval for $\mu$ or $\alpha$.

Here we take the estimation of $\mu$ as an example.

From (3.11) and (3.12), we have

$$\mu \geqslant -\frac{Y_n - Y_p}{R_n - R_p} = \mu_{floor} \tag{3.19}$$

for $R_n > R_p$. The interval of $\mu$ is given $[\mu_{floor}, \inf[$.

We can also set an upper bound. Let us consider the entire space. As $x_p$ is the best element, we should have $J_p \leqslant J_m$ where $R_m < R_p$ and $Y_m \geqslant Y_p$ (there is no element $x_m$ with $R_m < R_p$ and $Y_m < Y_p$, otherwise this $x_m$ is taken as $x_p$). Thus

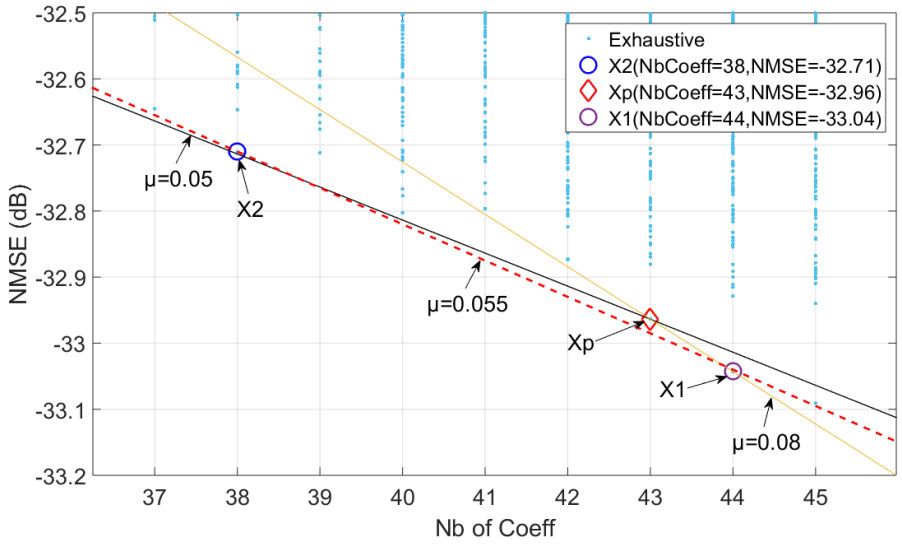



Figure 3.4: Determination of upper bound and lower bound of $\mu$ with a concave edge

we have

$$\mu \leqslant -\frac{Y_m - Y_p}{R_m - R_p} = \mu_{ceiling}. \tag{3.20}$$

In this way we can have a refined interval of $\mu$: $[\mu_{floor}, \mu_{ceiling}]$.

However as our search space is discrete, the edge of the blue cloud in Fig 3.3 can be concave. For instance, in Fig 3.4, if we define a such value of $\mu$ that $x_p$ and $x_1$ have the same merit value, we can plot a curve passing by these two points with slope equal to $\mu$. The lower bound $\mu_{floor} = 0.08$ is determined by $x_p$ and $x_1$, and the upper bound$\mu_{ceiling} = 0.05$ is determined by $x_p$ and $x_2$. From the inspection of Fig 3.4 we can see that $x_p$ lies in a concave region of the edge. This will result in an empty interval for $\mu$.

In this case we can observe that $\mu$ can be approximated to the slope of the dashed line passing by $x_1$ and $x_2$. Thus with the signal of the BS-PA, we take $0.055$ for the value of $\mu$.

For the multiplicative criterion, the upper bound is $1.4e^{-3}$, and the lower bound is $2.2e^{-3}$. In the same way, we take finally $1.5e^{-3}$ for the value of $\alpha$.

The GMP model space is represented into 3-D. The merit values are given in function of $\text{NMSE}_{\text{dB}}$ value, number of coefficients and the merit value in Fig 3.5.

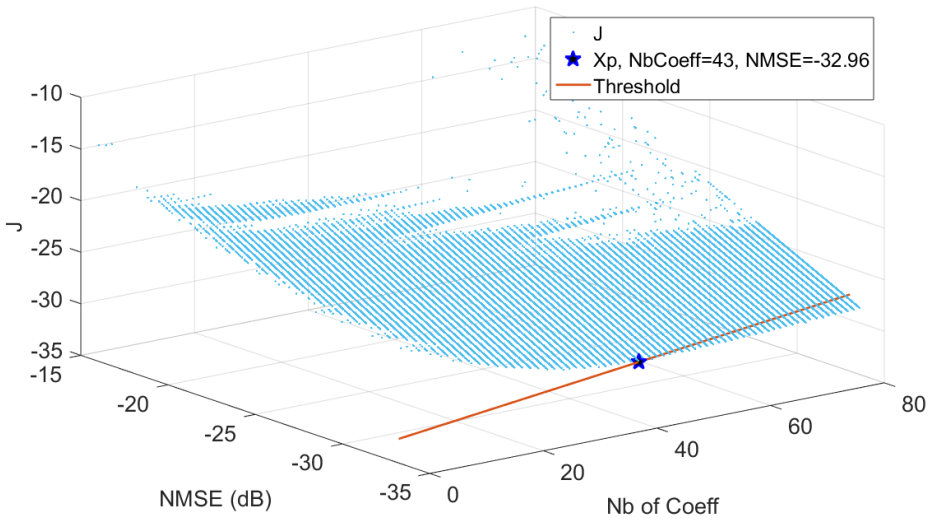



Figure 3.5: BS-PA: Exhaustive search results in function of $\text{NMSE}_{\text{dB}}$ and number of coefficients in 3D

With the signal of the BrC-PA, we get an interval of $\mu$ equal to $[0.0823, 0.0927]$ for the additive criterion. The mean value $\mu = 0.0875$ is taken.

For the multiplicative criterion, we get an interval of $\alpha$ equal to $[1.9e^{-3}, 2.2e^{-3}]$. The mean value $\alpha = 2e^{-3}$ is chosen.

We denote the element with the lowest merit value $J_{min}$ by $X_{min}$. The impact of different values of $\mu$ and $\alpha$ in the intervals on the algorithm performances and on $X_{min}$ is shown in Tables 3.1-3.4, where $SolNMSE$, $SolNb$, $NMSE_{X_{min}}$ and $Nb_{X_{min}}$ are the $\text{NMSE}_{\text{dB}}$ value and number of coefficients of the solution found by the algorithm and $X_{min}$ respectively. For base station PA, the solution and $X_{min}$ at the upper limit of interval of $\mu$ and $\alpha$ have more coefficients and also have better $\text{NMSE}_{\text{dB}}$. For broadcast PA, both the solution and $X_{min}$ are not sensitive to the values of $\mu$ and $\alpha$ in these intervals.

### 3.7.3 Test with Doherty PA for Base Station

#### Hill-Climbing heuristic using Off-line additive criterion (ACoff)

In 1-step search, we determine $\mu$ with the subspace of MP models before the GMP model search. The interval of $\mu$ is $[0.039, 0.070]$. If we choose the mean

Table 3.1: Additive criterion with different $\mu$ for BS-PA

$SolNMSE$: $\text{NMSE}_{\text{dB}}$ value of the solution found by the algorithm
$SolNb$: The number of coefficients of the solution found by the algorithm
$NMSE_{X_{min}}$: $\text{NMSE}_{\text{dB}}$ value of $X_{min}$
$Nb_{X_{min}}$: The number of coefficients of $X_{min}$

| $\mu$ | $SolNMSE$ | $SolNb$ | $NMSE_{X_{min}}$ | $Nb_{X_{min}}$ |
|---|---|---|---|---|
| 0.050 | -32.96 | 43 | -33.04 | 44 |
| 0.055 | -32.71 | 38 | -33.04 | 44 |
| 0.060 | -32.71 | 38 | -32.71 | 38 |
| 0.065 | -32.71 | 38 | -32.71 | 38 |
| 0.070 | -32.65 | 37 | -32.65 | 37 |
| 0.075 | -32.65 | 37 | -32.65 | 37 |
| 0.080 | -32.65 | 37 | -32.65 | 37 |

Table 3.2: Multiplicative criterion with different $\alpha$ for BS-PA

| $\alpha$ | $SolNMSE$ | $SolNb$ | $NMSE_{X_{min}}$ | $Nb_{X_{min}}$ |
|---|---|---|---|---|
| $1.4e^{-3}$ | -32.96 | 43 | -33.04 | 44 |
| $1.5e^{-3}$ | -32.70 | 38 | -33.04 | 44 |
| $1.6e^{-3}$ | -32.70 | 38 | -32.71 | 38 |
| $1.7e^{-3}$ | -32.70 | 38 | -32.71 | 38 |
| $1.8e^{-3}$ | -32.71 | 38 | -32.71 | 38 |
| $1.9e^{-3}$ | -32.71 | 38 | -32.71 | 38 |
| $2.0e^{-3}$ | -32.65 | 37 | -32.65 | 37 |
| $2.1e^{-3}$ | -32.65 | 37 | -32.65 | 37 |
| $2.2e^{-3}$ | -32.65 | 37 | -32.65 | 37 |

Table 3.3: Additive criterion with different $\mu$ for BrC-PA

| $\mu$ | $SolNMSE$ | $SolNb$ | $NMSE_{X_{min}}$ | $Nb_{X_{min}}$ |
|---|---|---|---|---|
| 0.082 | -38.50 | 18 | -40.64 | 29 |
| 0.084 | -38.50 | 18 | -40.56 | 28 |
| 0.086 | -38.50 | 18 | -40.56 | 28 |
| 0.087 | -38.50 | 18 | -40.56 | 28 |
| 0.088 | -38.50 | 18 | -40.56 | 28 |
| 0.090 | -38.50 | 18 | -40.56 | 28 |
| 0.092 | -38.50 | 18 | -40.56 | 28 |

Table 3.4: Multiplicative criterion with different $\alpha$ for BrC-PA

| $\alpha$ | $SolNMSE$ | $SolNb$ | $NMSE_{X_{min}}$ | $Nb_{X_{min}}$ |
|---|---|---|---|---|
| $1.9e^{-3}$ | -38.50 | 18 | -40.64 | 29 |
| $2.0e^{-3}$ | -38.50 | 18 | -40.56 | 28 |
| $2.1e^{-3}$ | -38.50 | 18 | -40.56 | 28 |
| $2.2e^{-3}$ | -38.50 | 18 | -40.37 | 26 |

value which is $0.055$, the final solution, $Sol$ in Fig 3.6, given by the algorithm is $(Y, R) = (-32.71, 38)$. The structure found is

$$\begin{aligned} \mathcal{K}_a &= 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b &= 4, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c &= 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1. \end{aligned} \tag{3.21}$$

In this search, there are totally 10821 different GMP models tested in 11 iterations. The execution time is 15.85 minutes. The searching iterations are indicated in Fig 3.6.

The iterations in function of $Y$ ($\text{NMSE}_{\text{dB}}$) and $R$ (Nb of Coeff) are depicted in Fig 3.6. The results of the exhaustive search are the background blue points (same for the following figures). The solutions of successive iterations are indicated by circles. The best element $X_{min}$ according to the criterion (3.1) with $\mu = 0.055$ has 44 coefficients with $\text{NMSE}_{\text{dB}}$ of -33.04 dB.

In Fig 3.6, the elements of the neighborhood tested in each iteration are shown in different colors. The point with an arrow "1" is the initial element $x_{0(1)}$. The points around with the same color are the neighbors of this element. The other black circles shows the best elements picked among their neighbors as the solution

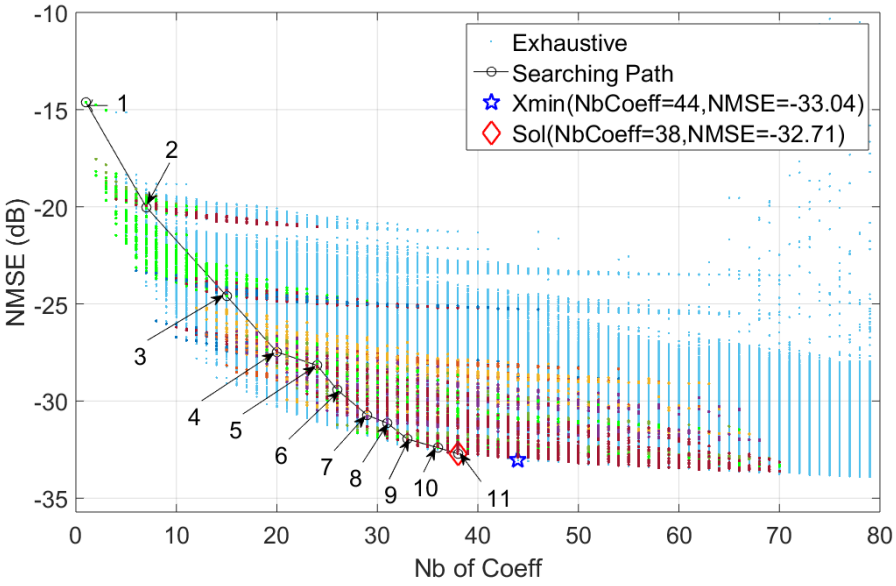
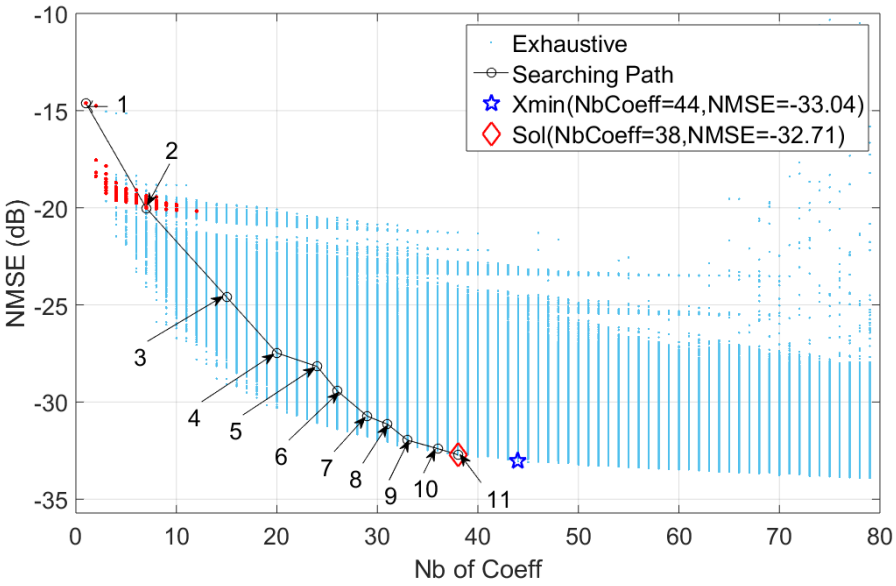



Figure 3.6: BS-PA: search path of Hill-Climbing heuristic with off-line additive criterion ($\mu = 0.055$)

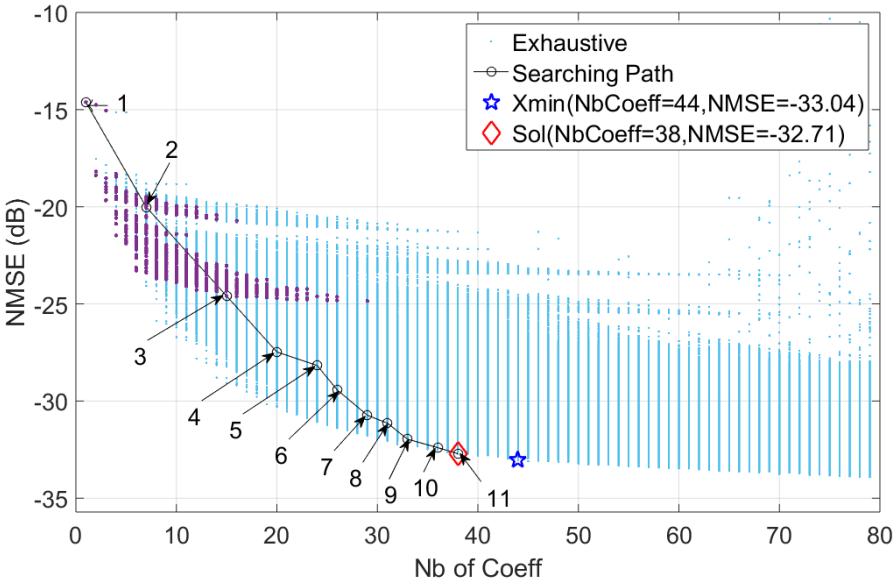
at each iteration.

The evolution of the algorithm is depicted in Fig 3.7 for the first two iterations. In Fig 3.7a, the red points are neighbors of $x_{0(1)}$, which are tested at the first iteration. Then the black circle "2" is selected as the solution $x_{s(1)}$. In Fig 3.7b, the purple points are neighbors of $x_{s(1)}$, which are tested at the second iteration, and the circle "3" is selected as $x_{s(2)}$. We can see that the search path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$. As each element will be tested at most only once, there is no overlap between the red points and the purple points.

With the same method to determine $\mu$, the two-step implementation (ACoff2) firstly searches only MP models, and then searches GMP models from the MP model solution. in the first step, there are 39 different MP models tested in 11 iterations during 3.6 seconds. The structure of the MP model with the best merit value found is $\mathcal{K}_a = 11$, $\mathcal{L}_a = 3$ with $\text{NMSE}_{\text{dB}}$ of -31.56 dB. And in the second step, starting from the element where $\mathcal{K}_a = 11$, $\mathcal{K}_b = 0$, $\mathcal{K}_c = 0$, $\mathcal{L}_a = 3$, $\mathcal{L}_b = 1$, $\mathcal{L}_c = 1$, $\mathcal{M}_b = 1$, $\mathcal{M}_c = 1$, there are 7037 different GMP model structures tested in 4 iterations during 12.64 minutes. The search path is depicted in Fig 3.8. The




(a) BS-PA: Neighborhood of the initial solution




(b) BS-PA: Neighborhood of the 2nd iteration

Figure 3.7: BS-PA: Neighborhoods demonstration of Hill-Climbing heuristic with additive criterion ($\mu = 0.055$)

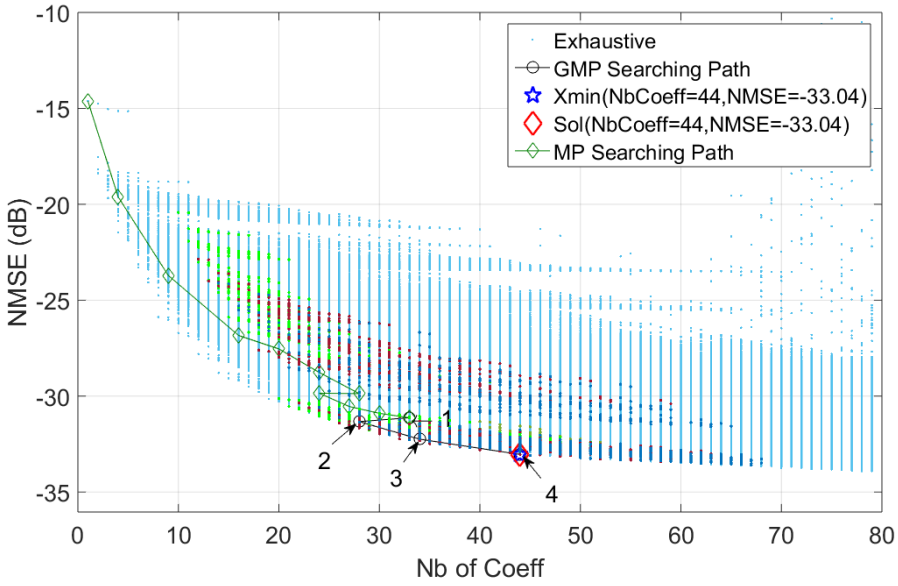



Figure 3.8: BS-PA: Two-step search path of Hill-Climbing heuristic with Off-line additive criterion ($\mu = 0.055$)

best structure $X_{min}$ found by the algorithm is

$$\begin{aligned} \mathcal{K}_a &= 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b &= 1, \mathcal{L}_b = 4, \mathcal{M}_b = 1 \\ \mathcal{K}_c &= 3, \mathcal{L}_c = 3, \mathcal{M}_c = 2. \end{aligned} \tag{3.22}$$

which has 44 coefficients and its $\text{NMSE}_{\text{dB}}$ value is -33.04 dB.

Using the additive criterion when $\mu$ determined with off-line computation, 2-step search can reduce the number of tests and execution time. As there are much less MP models than GMP models, by searching the best MP model, ACoff2 converges to $X_{min}$ with only a few tests. Here $X_{min}$ is evaluated with criterion (3.1) with $\mu = 0.055$. The GMP search step has only 4 iterations in Fig 3.8 instead of 11 iterations in Fig 3.6. The 2-step search avoids many tests in the zones which are not the neighborhood of $X_{min}$.

### Hill-Climbing heuristic using On-line additive criterion (ACon)

As mentioned before, in ACon, $\mu$ is determined with the subspace discovered at each iteration and its value is renewed according to (3.14). The evolution of the

Table 3.5: BS-PA: Evolution of $\mu$

| No. It | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mu$ | 0 | 0.242 | 0.176 | 0.163 | 0.148 | 0.116 |
| No. It | 7 | 8 | 9 | 10 | 11 | |
| $\mu$ | 0.107 | 0.086 | 0.075 | 0.058 | 0.055 | |

Table 3.6: BS-PA: Evolution of $\mu$ in Two Steps

| No. It | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\mu$ | 0 | 0.711 | 0.688 | 0.694 |

(a) MP search step

| No. It | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\mu$ | 0.829 | 0.139 | 0.112 | 0.105 |
| No. It | 5 | 6 | 7 | 8 |
| $\mu$ | 0.085 | 0.075 | 0.057 | 0.055 |

(b) GMP search step

value of $\mu$ is shown in Table 3.5. The value of $\mu$ converges towards the reference value 0.065 and becomes stable after 8 iterations.

In this search, the algorithm has 11 iterations and there are totally 8539 different GMP model structures tested during 12.78 minutes. The iterations in function of $Y$ ($\text{NMSE}_{\text{dB}}$) and $R$ (Nb of Coeff) are depicted in Fig 3.9. The solutions of iterations are indicated by circles with annotations. We can see that the search path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$. Here $X_{min}$, the same as (3.22), is evaluated based on the latest value of $\mu$ which is 0.055. According to all these ($Y$, $R$) couples, we found the solution the same as (3.21) which has 38 coefficients and $\text{NMSE}_{\text{dB}}$ of $-32.71$ dB.

In 2-step search (ACon2), with the same method to estimate $\mu$, we can reach the best element in 12 iterations. The evolution of the value of $\mu$ is shown in Table 3.6. With the final value $\mu = 0.055$, $X_{min}$ given by the criterion (3.2) has 44 coefficients with $\text{NMSE}_{\text{dB}}$ equal to -33.04 dB, and its structure is the same as (3.22).

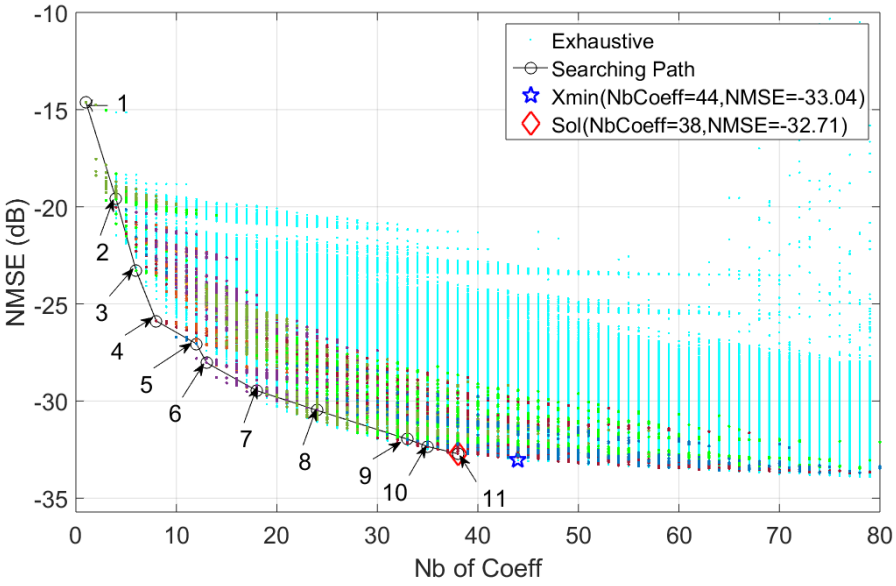



Figure 3.9: BS-PA: On-line additive criterion in function of $\text{NMSE}_{\text{dB}}$ and number of coefficients ($\mu = 0.055$)

In the first step, there are 15 different MP models tested in 4 iterations during 0.649 seconds. The structure of the MP model with the best merit value found is $\mathcal{K}_a = 4$, $\mathcal{L}_a = 2$ with $\text{NMSE}_{\text{dB}}$ of -25.88 dB. In the second step, starting from the element where $\mathcal{K}_a = 4$, $\mathcal{K}_b = 0$, $\mathcal{K}_c = 0$, $\mathcal{L}_a = 2$, $\mathcal{L}_b = 1$, $\mathcal{L}_c = 1$, $\mathcal{M}_b = 1$, $\mathcal{M}_c = 1$, there are 8155 different GMP model structures tested in 8 iterations during 12.58 minutes. The search path is depicted in Fig 3.10. The blue points are again the exhaustive results, and the circles with annotations are the solutions of iterations in the Hill-Climbing heuristic. The best structure found by the algorithm is also (3.21).

Comparing ACon2 with ACon, we can see that, 2-step search does not reduce the number of tests and execution time in on-line computation as it did in off-line computation. That is due to the inaccurate determination of $\mu$ during the first step. In MP model search, we tested only 15 MP models in 4 iterations. In other words, less than 4 models in average are used to determine $\mu$ at each iteration. Determining the value of $\mu$ with these small size sampling sometimes get a result far from its final converged value, which can be seen in Table 3.6. Thus in this case the first step does not improve the algorithm.

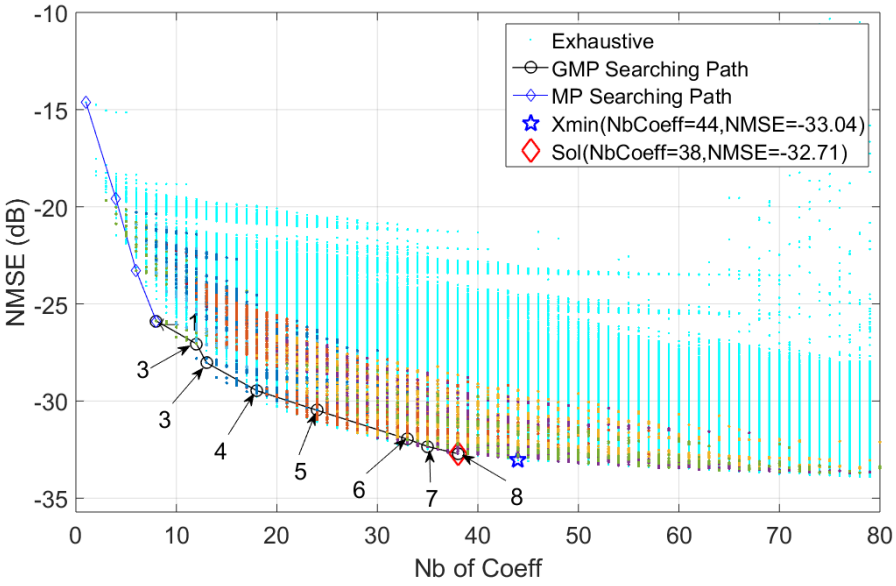



Figure 3.10: BS-PA: Two-step search path of Hill-Climbing heuristic with On-line additive criterion ($\mu = 0.055$)

Furthermore with the on-line computation, we can see that the search path follows the envelop of the blue points cloud, which can reduce the computation complexity. Further discussions are presented in Section 3.7.5.

### Hill-Climbing heuristic using Off-line multiplicative criterion (MCoff)

Using the MP models subspace, the interval for $\alpha$ is $[0.0012, 0.0021]$ and the mean value $1.7e^{-3}$ has been chosen. With this $\alpha$, the best element $X_{min}$ given by the criterion is the same as (3.21).

In 1-step search, the algorithm has 11 iterations and there are totally 19598 different GMP model structures tested. The iterations in function of $Y$ ($\text{NMSE}_{\text{dB}}$) and $R$ (Nb of Coeff) are depicted in Fig 3.11. The solutions of iterations are indicated by circles with annotations. We can see that the search path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$ which is (3.21). The execution time is 29.89 minutes.

In two-step search (MCoff2), in the first step, there are 39 different MP models tested in 11 iterations during 3.6 seconds. The structure of the MP model with the best merit value found is $\mathcal{K}_a = 11$, $\mathcal{L}_a = 3$ with $\text{NMSE}_{\text{dB}}$ of -31.13 dB. In

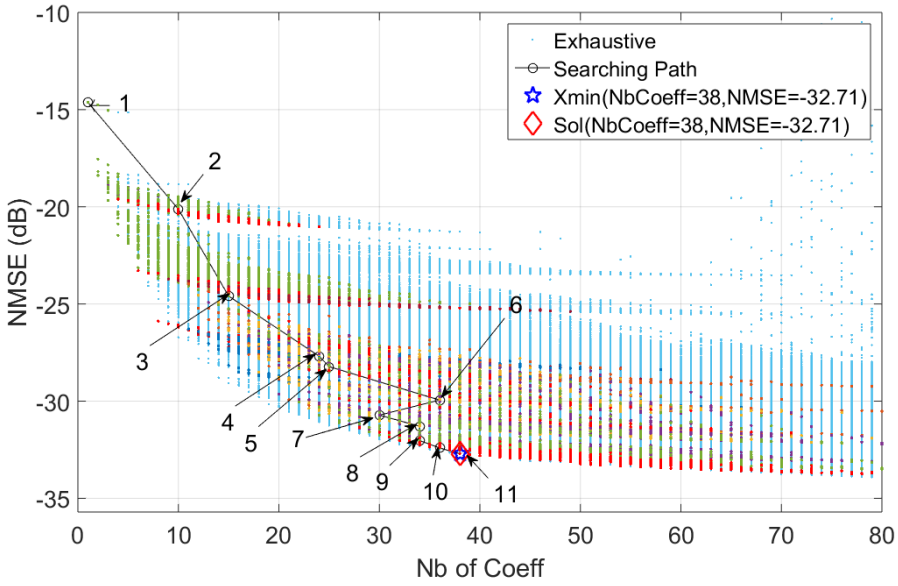



Figure 3.11: BS-PA: Off-line multiplicative criterion in function of $\text{NMSE}_{\text{dB}}$ and number of coefficients ($\alpha = 1.7e^{-3}$)

the second step, starting from the element where $\mathcal{K}_a = 11$, $\mathcal{K}_b = 0$, $\mathcal{K}_c = 0$, $\mathcal{L}_a = 3$, $\mathcal{L}_b = 1$, $\mathcal{L}_c = 1$, $\mathcal{M}_b = 1$, $\mathcal{M}_c = 1$, there are 7364 different GMP model structures tested with 4 iterations during 14.20 minutes. The search path is depicted in Fig 3.12. The solution found by the algorithm is the same as (3.22) which has 44 coefficients and its $\text{NMSE}_{\text{dB}}$ value is -33.04 dB.

The value of $\alpha$ estimated by off-line computation, $1.7e^{-3}$, is close to the reference value, $1.8e^{-3}$. Comparing with MCoff, we can see that MCoff2 reduces the number of tests and execution time with a solution close to the optimum.

### Hill-Climbing heuristic using On-line multiplicative criterion (MCon)

Using the subspace composed of the models tested at each iteration, we can determine $\alpha$ dynamically.

In 1-step search, the algorithm has 13 iterations and there are totally 12692 different GMP model structures tested during 17.12 minutes. The evolution of the value of $\alpha$ is shown in Table 3.7.

$X_{min}$ given by the criterion (3.2) with $\alpha = 0.0016$ is the same as (3.22). The iterations in function of $Y$ ($\text{NMSE}_{\text{dB}}$) and $R$ (Nb of Coeff) are depicted in Fig 3.13.

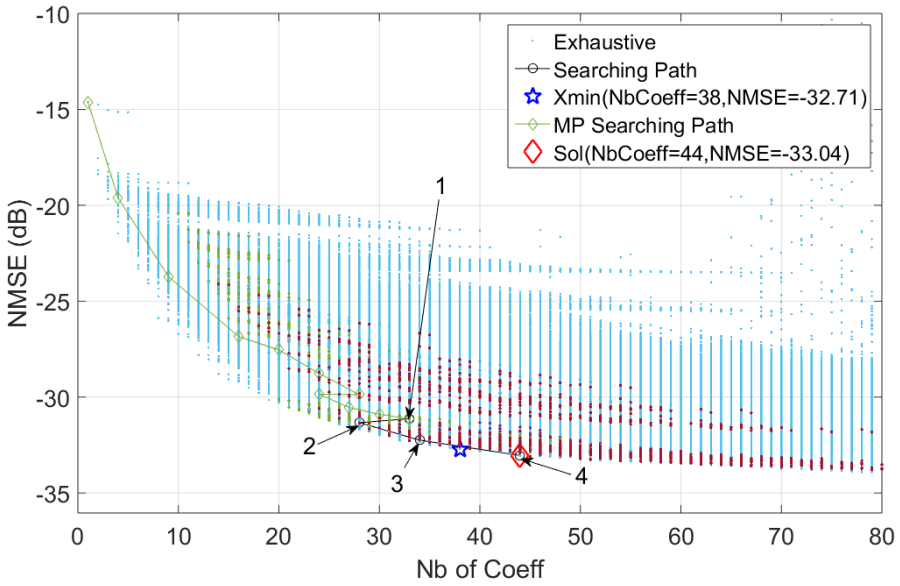



Figure 3.12: BS-PA: Two-step search path of Hill-Climbing heuristic with Off-line multiplicative criterion ($\alpha = 1.7e^{-3}$)

Table 3.7: BS-PA: Evolution of $\alpha$

| No. It | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\alpha$ | 0 | 11.7 | 0.0079 | 0.0064 | 0.0054 |
| No. It | 6 | 7 | 8 | 9 | 10 |
| $\alpha$ | 0.0038 | 0.0034 | 0.0027 | 0.0022 | 0.0024 |
| No. It | 11 | 12 | 13 | | |
| $\alpha$ | 0.0019 | 0.0017 | 0.0016 | | |

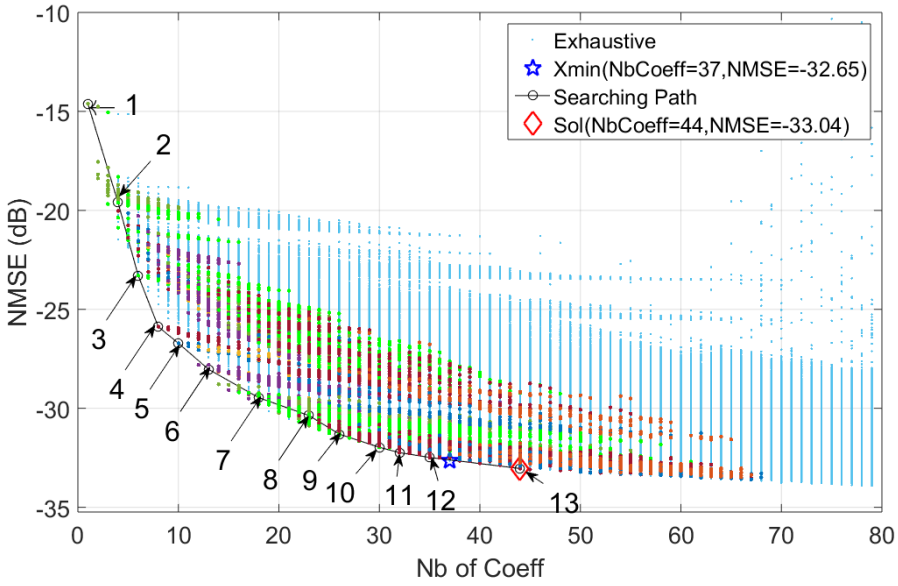



Figure 3.13: BS-PA: On-line multiplicative criterion in function of $\text{NMSE}_{\text{dB}}$ and number of coefficients ($\alpha = 1.6e^{-3}$)

According to all these ($Y$, $R$) couples, the algorithm found the best structure with 44 coefficients having $\text{NMSE}_{\text{dB}}$ of $-33.04$ dB as (3.22).

In two-step search (MCon2), the best element is reached in 12 iterations. In the first step, there are 15 different MP models tested in 4 iterations during 1.7138 seconds. The evolution of the value of $\alpha$ is shown in Table 3.8. $X_{min}$ given by the criterion (3.1) with $\alpha = 1.6e^{-3}$ is the same with (3.22).

The structure of the MP model with the best merit value found is $\mathcal{K}_a = 4$, $\mathcal{L}_a = 2$ with $\text{NMSE}_{\text{dB}}$ of -25.88 dB. In the second step, starting from the element where $\mathcal{K}_a = 4$, $\mathcal{K}_b = 0$, $\mathcal{K}_c = 0$, $\mathcal{L}_a = 2$, $\mathcal{L}_b = 1$, $\mathcal{L}_c = 1$, $\mathcal{M}_b = 1$, $\mathcal{M}_c = 1$, there are 12692 different GMP model structures tested in 10 iterations during 18.31 minutes. The search path is depicted in Fig 3.14. The best structure found by the algorithm is the same as (3.22) which has 44 coefficients and its $\text{NMSE}_{\text{dB}}$ value is -33.04 dB.

The first step of the algorithm does not help reducing the execution time because the determination of $\alpha$ decays with very small number of MP models. The final result is the same with MCon.

Table 3.8: BS-PA: Evolution of $\alpha$ in Two Steps

| No. It | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\alpha$ | 0 | 0.0169 | 0.0219 | 0.0220 |

(a) MP search step

| No. It | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\alpha$ | 0.0262 | 0.0048 | 0.0038 | 0.0034 | 0.0027 |
| No. It | 6 | 7 | 8 | 9 | 10 |
| $\alpha$ | 0.0022 | 0.0024 | 0.0019 | 0.0017 | 0.0016 |

(b) GMP search step

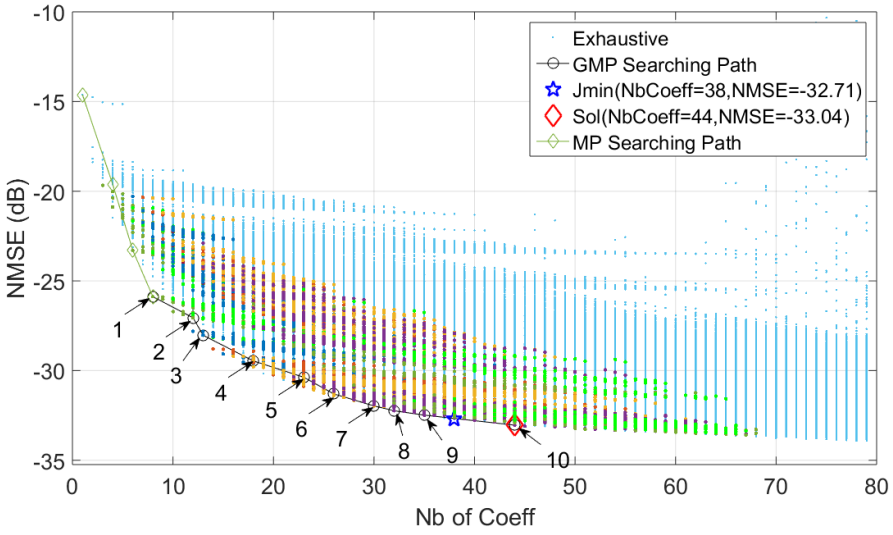



Figure 3.14: BS-PA: Two-step search path of Hill-Climbing heuristic with On-line multiplicative criterion ($\alpha = 1.6e^{-3}$)

### 3.7.4 Test with Doherty PA for Broadcast

**Hill-Climbing heuristic using Off-line additive criterion (ACoff)**

The estimated interval of $\mu$ is $[0.163, 0.263]$ with MP model subspace, and we take the mean $\mu = 0.213$ which is outside the reference interval. In this case we found that the determined value of $\mu$ is much different from the reference value $\mu = 0.0875$ in Section 3.7.2. As the sample size of off-line computation is not large enough, the estimation is not always accurate.

In 1-step search, we can find a solution in 5 iterations. There are totally 4608 GMP model structures tested. The search path is depicted in Fig 3.15. $X_{min}$ given by the criterion (3.1) with $\mu = 0.213$ has 15 coefficients and $\text{NMSE}_{\text{dB}}$ value of -38.7473 dB, and its structure is

$$\begin{aligned} \mathcal{K}_a =& 8, \mathcal{L}_a = 1 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 2, \mathcal{L}_c = 3, \mathcal{M}_c = 1. \end{aligned} \tag{3.23}$$

The search path ends at the point with a distance of $(3, -1.22dB)$ from to $X_{min}$ which is the best GMP model structure according to the additive criterion. The parameters of the final solution are:

$$\begin{aligned} \mathcal{K}_a =& 5, \mathcal{L}_a = 1 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 2, \mathcal{L}_c = 3, \mathcal{M}_c = 1 \end{aligned} \tag{3.24}$$

with $\text{NMSE}_{\text{dB}}$ of -37.52 dB and 12 coefficients. The execution time is 4.50 minutes.

Using 2-step approach (ACoff2), first we find the MP model ($\mathcal{K}_a = 4, \mathcal{L}_a = 2$) with $\text{NMSE}_{\text{dB}}$ of -32.82 dB. In this step, 14 MP model structures are tested in 4 iterations during 1.28 seconds. From this MP model, we can find a GMP model structure as the final solution with 3 iterations, and 3606 GMP model structures are tested during 3.47 minutes. The final solution is the same as (3.24) with $\text{NMSE}_{\text{dB}}$ of -37.52 dB and 12 coefficients. In Fig 3.16 we can see that the search path ends at the point with a distance of $(3, -1.22dB)$ from the star $X_{min}$ which is the same as (3.23).

**Hill-Climbing heuristic using On-line additive criterion (ACon)**

In 1-step search, with the additive criterion, we can find a solution in 5 iterations. There are totally 6168 GMP model structures tested during 6.41 minutes. The

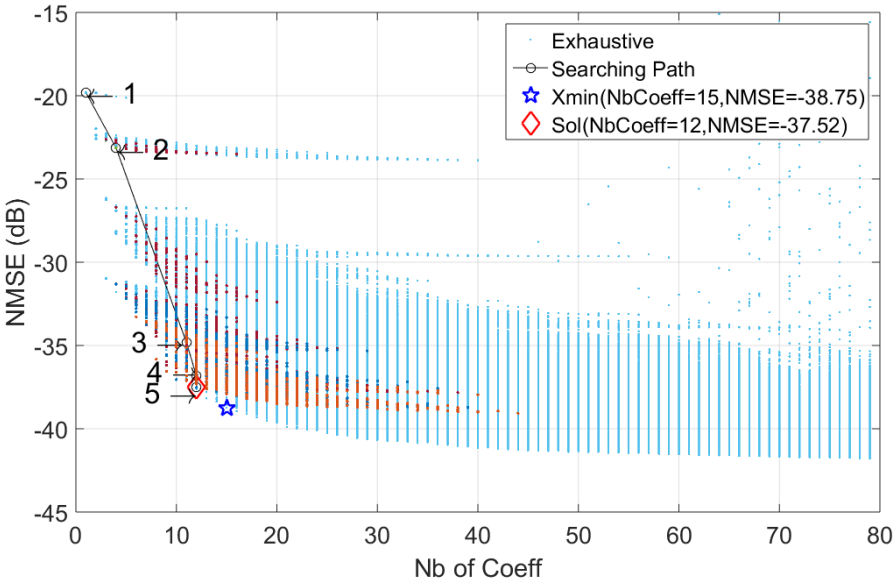



Figure 3.15: BrC-PA: search path of Hill-Climbing heuristic with Off-line additive criterion ($\mu = 0.213$)

search path is depicted in Fig 3.17. The final solution which is the same as

$$\begin{aligned} \mathcal{K}_a =& 5, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 3, \mathcal{L}_c = 2, \mathcal{M}_c = 1 \end{aligned} \tag{3.25}$$

which has 17 coefficients and $\text{NMSE}_{\text{dB}}$ of -38.20 dB.

It has a distance of $(4, -1.69dB)$ to $X_{min}$ which is the best GMP model structure according to the additive criterion. Here $X_{min}$ evaluated based on the latest value $\mu = 0.1416$ has 21 coefficients and $\text{NMSE}_{\text{dB}}$ of -39.89 dB. The structure of $X_{min}$ is

$$\begin{aligned} \mathcal{K}_a =& 4, \mathcal{L}_a = 1 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 8, \mathcal{L}_c = 2, \mathcal{M}_c = 1. \end{aligned} \tag{3.26}$$

The evolution of the value of $\mu$ is shown in Table 3.9.

Using 2-step approach (ACon2), first we find the MP model ($\mathcal{K}_a = 5, \mathcal{L}_a = 2$) with $\text{NMSE}_{\text{dB}}$ of -33.16 dB. In this step, 14 MP model structures are tested in 4

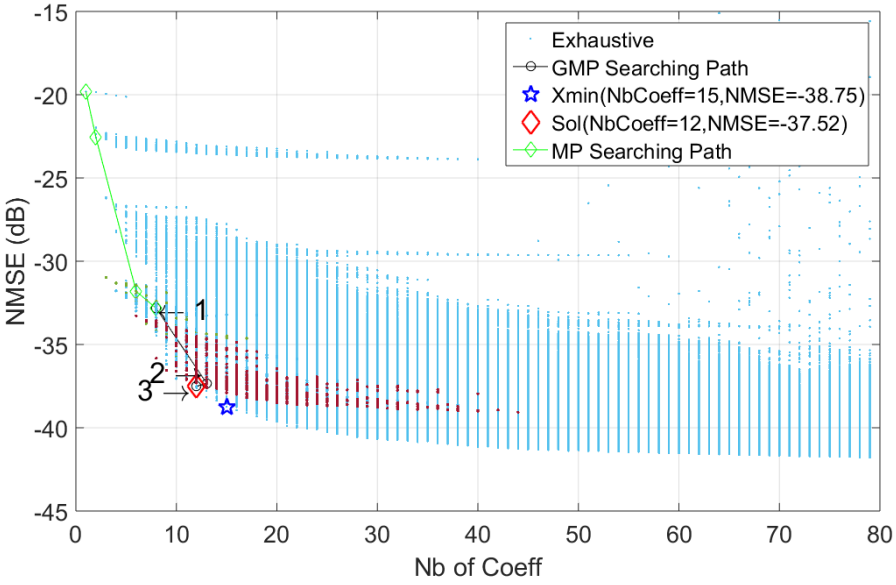



Figure 3.16: BrC-PA: Two-step search path of Hill-Climbing heuristic with Off-line additive criterion ($\mu = 0.213$)

Table 3.9: BrC-PA: Evolution of $\mu$

| No. Iter | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\mu$ | 0 | 0.0686 | 0.2150 | 0.1954 |
| No. Iter | 5 | 6 | 7 | |
| $\mu$ | 0.1933 | 0.1923 | 0.1416 | |

Table 3.10: BrC-PA: Evolution of $\mu$ in Two Steps

| No. It | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\mu$ | 0 | 0.0775 | 0.1223 | 0.1309 |

(a) MP search step

| No. It | 1 | 2 | 3 |
|---|---|---|---|
| $\mu$ | 0.0883 | 0.2781 | 0.2199 |

(b) GMP search step

iterations during 1.28 seconds. From this MP model, a GMP model structure as the final solution is found with 3 iterations, and 2856 GMP model structures have been tested during 2.91 minutes. This final solution is the same as (3.24) with 12 coefficients and $\text{NMSE}_{\text{dB}}$ of -37.52 dB.

In Fig 3.18 we can see the solution of the search path has a distance of $(9, -2.48dB)$ to the star which represents $X_{min}$. Here $X_{min}$ evaluated based on the latest value $\mu = 0.2199$ is the same as (3.23).

The evolution of the value of $\mu$ is shown in Table 3.10.

**Hill-Climbing heuristic using Off-line multiplicative criterion (MCoff)**

With the Off-line multiplicative criterion, the interval of $\alpha$ is $[0.0077, 0.0047]$, and we take the mean $\alpha = 6.2e^{-3}$.

In 1-step search, the solution is found in 5 iterations. There are totally 3056 GMP model structures tested. The search path is depicted in Fig 3.19. This final solution as

$$\begin{aligned} \mathcal{K}_a &= 5, \mathcal{L}_a = 1 \\ \mathcal{K}_b &= 1, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c &= 2, \mathcal{L}_c = 2, \mathcal{M}_c = 1 \end{aligned} \tag{3.27}$$

with $\text{NMSE}_{\text{dB}}$ of -37.04 dB and 10 coefficients. It approaches $X_{min}$ which is the best GMP model structure according to the multiplicative criterion with $\alpha = 6.2e^{-3}$. Here $X_{min}$ is the same as (3.23). The execution time is 2.66 minutes.

Using 2-step approach (MCoff2), first we find the MP model ($\mathcal{K}_a = 4, \mathcal{L}_a = 2$) with $\text{NMSE}_{\text{dB}}$ of -32.82 dB. In this step, 14 MP model structures are tested during

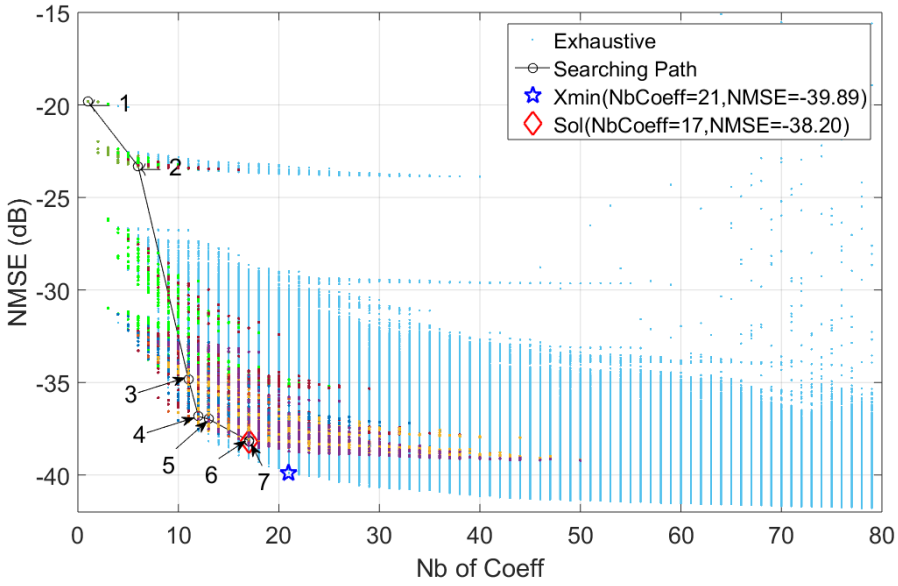



Figure 3.17: BrC-PA: search path of Hill-Climbing heuristic with On-line additive criterion ($\mu = 0.1416$)

2.35 seconds. From this MP model, we can find a GMP model structure as the final solution with 3 iterations, and 2568 GMP model structures are tested during 2.66 minutes. The final solution is the same as (3.27) and $X_{min}$ is the same as (3.23). In Fig 3.20 we can see the search path and it ends at the point with a distance of $(3, -1.22dB)$ from the star which represents $X_{min}$ which is evaluated by the multiplicative criterion with $\alpha = 6.2e^{-3}$.

**Hill-Climbing heuristic using On-line multiplicative criterion (MCon)**

In 1-step search, with the multiplicative criterion, we can find a solution in 5 iterations. There are totally 3820 GMP model structures tested during 3.41 minutes. The search path is depicted in Fig 3.21. The final solution is the same as (3.24). Here $X_{min}$ evaluated based on the latest value of $\alpha$ which is 0.0050 is the same as (3.23).

The evolution of the value of $\alpha$ is shown in Table 3.11.

Using 2-step approach (MCon2), first we find the MP model ($\mathcal{K}_a = 4, \mathcal{L}_a = 2$) with $\text{NMSE}_{\text{dB}}$ of -32.83 dB. In this step, 15 MP model structures are tested in 4 iterations during 1.79 seconds. From this MP model, we can find a GMP model

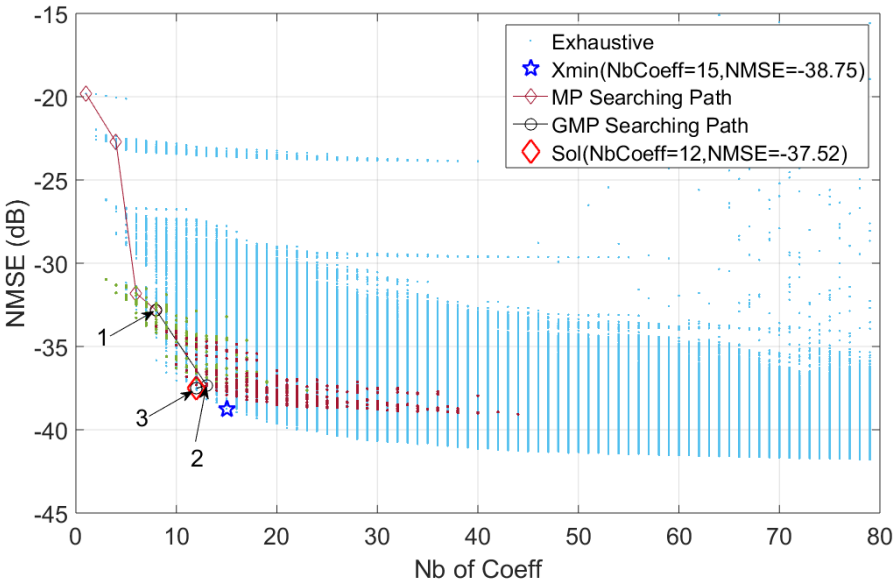



Figure 3.18: BrC-PA: Two-step search path of Hill-Climbing heuristic with On-line additive criterion ($\mu = 0.2025$)

Table 3.11: BrC-PA: Evolution of $\alpha$

| No. It | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\alpha$ | 0 | 0.0030 | 0.0063 | 0.0053 | 0.0050 |

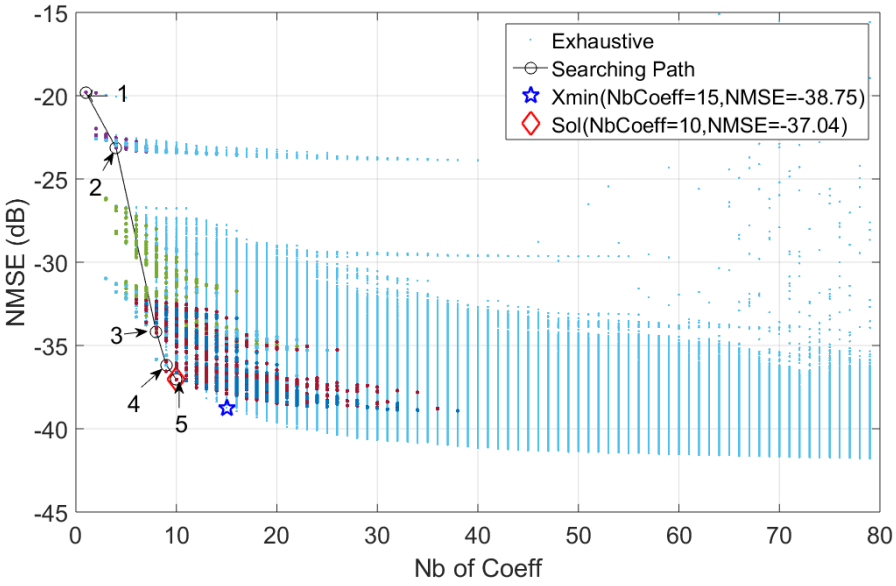



Figure 3.19: BrC-PA: search path of Hill-Climbing heuristic with Off-line multiplicative criterion ($\alpha = 6.2e^{-3}$)

structure as the final solution with 4 iterations, and 3000 GMP model structures are tested during 3.09 minutes. The final solution is the same as (3.24).

In Fig 3.22 we can see the search path which ends at the point with a distance of $(3, -1.22dB)$ from the star which represents $X_{min}$. Here $X_{min}$ evaluated based on the latest value of $\alpha = 0.0038$ is the same as (3.23).

The evolution of the value of $\alpha$ is shown in Table 3.12.

### 3.7.5 Discussion

For the exhaustive results which are depicted by the light blue point cloud in Fig 3.5 - Fig 3.22, the left and lower edge of the point cloud is defined as its envelop. The slope of this envelop is steep on the left edge and flat on the lower edge. The connection of these two edges is the turning area where the slope decreases prominently. For broadcast PA, e.g.

The normalized execution times show the comparative complexities of different search procedures. The total execution time of ACoff (search plus off-line determination of $\mu$, which lasts 15.85 minutes in case of base station PA and 4.50

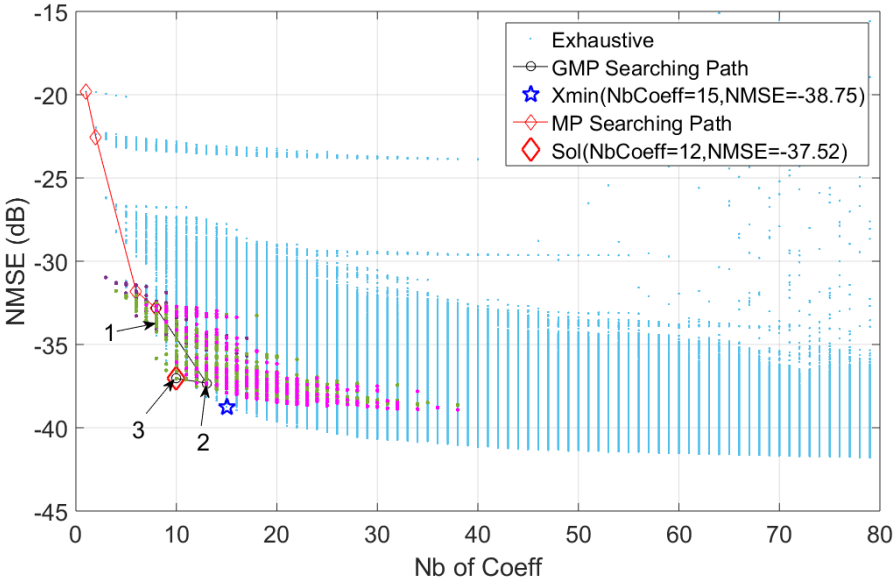



Figure 3.20: BrC-PA: Two-step search path of Hill-Climbing heuristic with Off-line multiplicative criterion ($\alpha = 6.2e^{-3}$)

Table 3.12: BrC-PA: Evolution of $\alpha$ in Two Steps

| No. It | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\alpha$ | 0 | 0.0017 | 0.0035 | 0.0038 |

(a) MP search step

| No. It | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\alpha$ | 0.0062 | 0.0060 | 0.0051 | 0.0038 |

(b) GMP search step

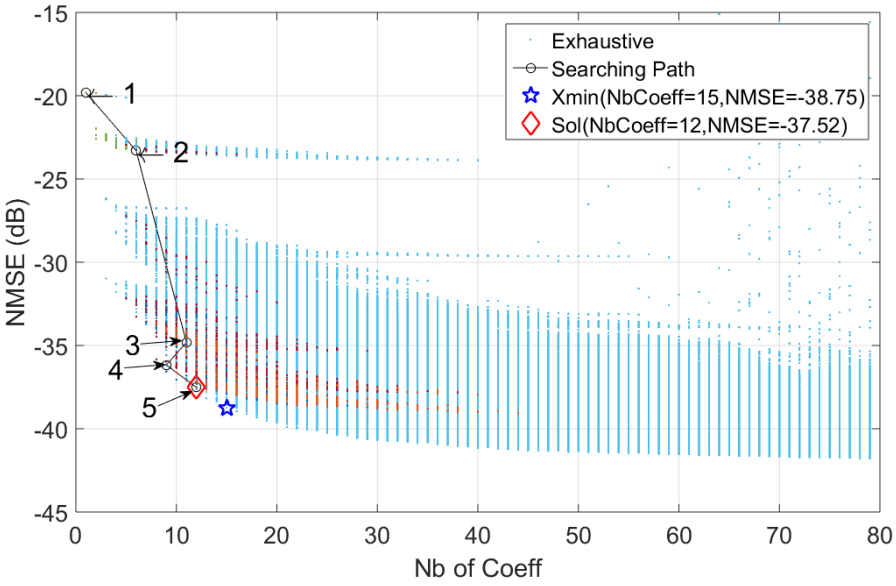



Figure 3.21: BrC-PA: search path of Hill-Climbing heuristic with On-line multiplicative criterion ($\alpha = 5e^{-3}$)

minutes in case of broadcast PA) is taken as the unit time. The complexity is represented by the execution time and the number of tested model which are provided in the Tables summarizing the performances of the different implementations.

The exhaustive search covers all possible models with $R < 80$. So $X_{min}$ can be regarded as the global optimum. Though using Hill-Climbing heuristic can not always reach $X_{min}$, as sometimes it seems stuck at a local optimum when the search path does not reach $X_{min}$ or its neighbors, e.g. Fig 3.19 and Fig 3.20, it is still a reliable method to size the GMP models. The results show that the search path starts from the simplest model straightly towards the global optimum. And we can observe that the found solution is most of the time very close to the global optimum and located on the envelop.

Both the additive criterion and the multiplicative criterion achieve a good performance. For the signal of base station PA, the search with additive criterion takes less execution time than the multiplicative criterion. For the signal of broadcast PA, the multiplicative criterion has advantage on execution time.

According to the results, the velocity of algorithm convergence depends mainly on two factors: the number of tested models and the complexity of tested models.

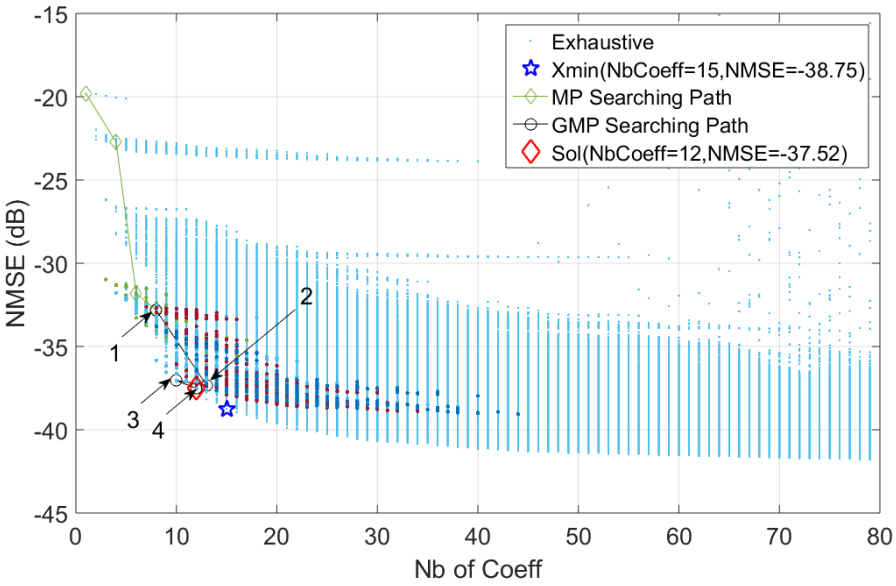



Figure 3.22: BrC-PA: Two-step search path of Hill-Climbing heuristic with On-line multiplicative criterion ($\alpha = 3.8e^{-3}$)

Table 3.13: Comparison of GMP Results of Different Searches with BS-PA

| | ACoff | ACoff2 | ACon | ACon2 |
|---|---|---|---|---|
| $NMSE_{dB}$ | -32.71 | -33.04 | -32.71 | -32.71 |
| NbCoeff | 38 | 44 | 38 | 38 |
| Nb of Test | 10821 | 39+7037 | 8539 | 15+8155 |
| Nb of Iter | 11 | 11+4 | 11 | 4+8 |
| Ex. Time | 1 | 0.80 | 0.81 | 0.79 |
| | MCoff | MCoff2 | MCon | MCon2 |
| $NMSE_{dB}$ | -32.71 | -33.04 | -33.04 | -33.04 |
| NbCoeff | 38 | 44 | 44 | 44 |
| Nb of Test | 19598 | 39+7364 | 12692 | 15+12692 |
| Nb of Iter | 11 | 11+4 | 13 | 4+10 |
| Ex. Time | 1.89 | 0.90 | 1.08 | 1.16 |

Table 3.14: Comparison of GMP Results of Different Searches with BrC-PA

| | ACoff | ACoff2 | ACon | ACon2 |
|---|---|---|---|---|
| $NMSE_{dB}$ | -37.52 | -38.75 | -38.20 | -37.52 |
| NbCoeff | 12 | 15 | 17 | 12 |
| Nb of Test | 4608 | 14+3606 | 6168 | 14+2856 |
| Nb of Iter | 5 | 4+3 | 7 | 4+3 |
| Ex. Time | 1 | 0.77 | 1.42 | 0.65 |
| | MCoff | MCoff2 | MCon | MCon2 |
| $NMSE_{dB}$ | -37.04 | -37.04 | -37.52 | -37.52 |
| NbCoeff | 10 | 10 | 12 | 12 |
| Nb of Test | 3056 | 18+2568 | 3820 | 14+3000 |
| Nb of Iter | 5 | 5+3 | 5 | 4+4 |
| Ex. Time | 0.59 | 0.56 | 0.76 | 0.69 |

From Table 3.13 and Table 3.14, we can see that the number of tested models has impact on the execution time, but not in a proportional way. For example in Table 3.13 ACoff has 25406 GMP models tested during 39.91 minutes while ACoff2 has 10130 GMP models tested during 19.99 minutes. The average execution time per model of ACoff2 (which is $0.1184$ s/model) is greater than that of ACoff (which is $0.0943$ s/model). That is because ACoff2 starts to search GMP models from an element which has high order of nonlinearity and/or deep memory depth and the neighbors of this element have very large number of basis functions.

The points on the envelop correspond to the GMP models with the simplest structure among those of the same $NMSE_{dB}$ value. When the path follows the envelop, the tested models have less complexity compared with models which are not on the envelop. When it follows the envelop as in Fig 3.19 than having zigzag shapes, e.g. in Fig 3.11 or in Fig 3.20, the search path has less average execution time.

The determination of weighting coefficients is a challenge in construction of merit functions. According to (3.11), the final solution should have the least number of coefficients while its $NMSE_{dB}$ is better than $bY_{min}$. With the estimation methods proposed in this chapter, we can restrict the weighting coefficient to a given range. On-line computation allows to refine the value of weighting coefficient as it has more samples for the estimation.

With the signal of the broadcast PA, the final solutions found by the algorithms are not located exactly at but very close to the points of $X_{min}$. The search

path stops at the local optimum after approaching the zone where the values of $NMSE_{dB}$ get stable.

The broadcast PA can be well modeled by a GMP model with low nonlinearity order and memory depth ($NMSE_{dB}$ <-37 dB). Thus we can see that the envelop falls down very fast when $R < 10$ in Fig 3.15-3.22. When $R > 20$ we can see that the envelop is stable in terms of $NMSE_{dB}$. The area $10 < R < 20$ is the turning area. Once a model in this area is reached, we need to add much more coefficients to have the same improvement of modeling accuracy.

For the algorithms with weighting coefficients determined with off-line method, 2-step search always reduces the number of tests and execution time for both PAs. The search for the best MP model is faster because an MP model has much fewer neighbors than a GMP model. Starting from the linear memoryless model, MP model search reaches a region towards the best element rapidly, e.g. Fig 3.8.

For the algorithms with weighting coefficients determined with on-line computation, the advantage of 2-step search is very limited. The tested models in the first step are too few to assure the determined weighting coefficient values close to the reference value. Therefore the searching path of this step is not beneficial.

In the next section, we present the results obtained with different pruned neighborhoods. Only MCoff criterion is applied to make a comparison with the conventional algorithm. The improvements on execution time is obvious and the final searched results are always kept in the region of global optimum.

### 3.7.6 Results of HC with pruned neighborhoods

#### HC with constraint on Number of Coefficients

In this test, for the actual solution $x_{0(q)}$ at the $q$-th iteration, only the elements $x_{i(q)}$ with $|R_{i(q)} - R_{0(q)}| \leqslant d$ are tested.

The search result of HC using MCoff criterion ($\alpha = 1.7e^{-3}$) with constraint on number of coefficients tested on BS-PA are shown in this section.

In 1-step search, there are totally 3835 different GMP model structures tested. The number of coefficients variation limit starts from $d = 2$ and ends with $d = 3$. The elements in function of $Y$ ($NMSE_{dB}$) and $R$ (Nb of Coeff) are depicted in Fig 3.23. The searching steps are also indicated in this figure. We can see that the searching path converges directly towards the low criterion merit value region. The best structure is given as $X_{min}$ in this figure. The solution is the same as (3.21) with 38 coefficients having NMSE of -32.71 dB. The result of the algorithm shown in Fig 3.23 successfully finds the best element.

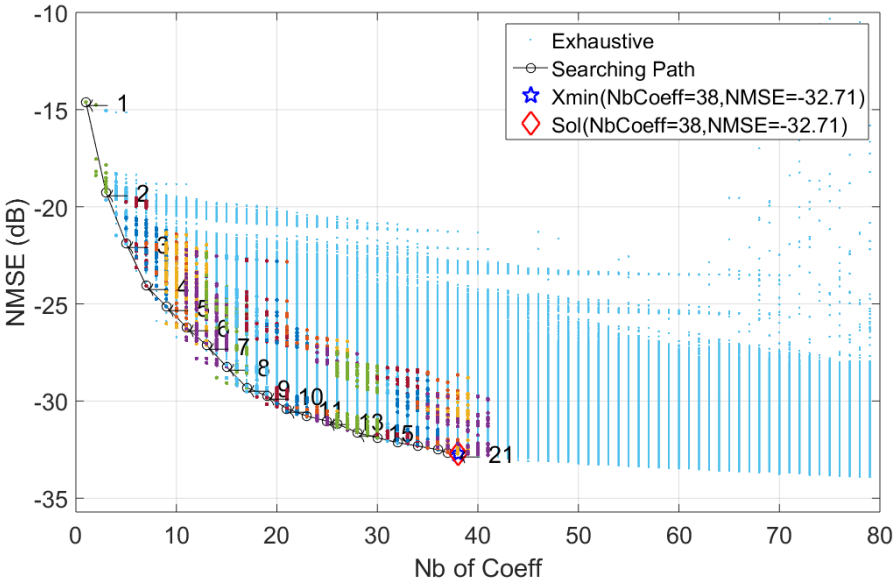



Figure 3.23: Seaching path of Hill-Climbing with constraint on Number of Coefficients

In the criterion, the value of the constant $\alpha$ is an influential factor which affects the merit function and may also have an impact on decision of the searching path.

There are totally 21 steps during 4.71 minutes and we can see that the searching path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$ in this figure.

In two-step search, first we find out the MP model minimizing the merit function (3.2). Then we start from this MP model, to find the best GMP model. In this search, we can reach the best element in 7 iterations. In the first step, there are 39 different MP models tested. The structure of the MP model with the best NMSE value found is $\mathcal{K}_a = 11$, $\mathcal{L}_a = 3$ with -31.13 dB NMSE. The searching path in map of NMSE vs number of coefficients is depicted in Fig 3.24. The blue points are exhaustive results, and the colorful points are tested by the Hill-Climbing heuristic. There are 959 different GMP model structures tested during 1.77 minutes. The best structure found by the algorithm is the same as (3.21) with 38 coefficients and its NMSE value is -32.71 dB. The number of coefficients variation limit starts from $d = 2$ and ends with $d = 3$.

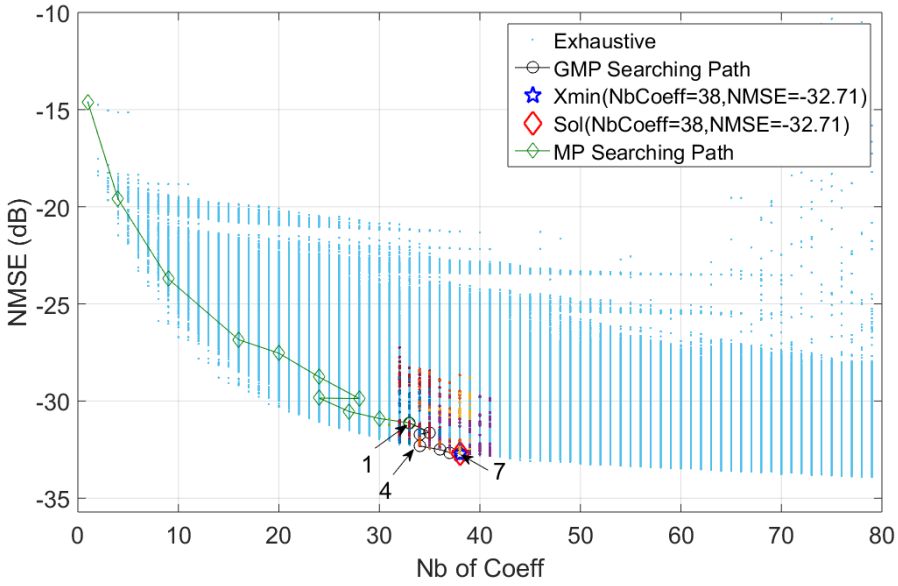



Figure 3.24: Two-step seaching path of Hill-Climbing with constraint on Number of Coefficients

### HC with jumping on Number of Coefficients

In this test, for the actual solution $x_{0(q)}$ at the $q$-th iteration, only the elements $x_{i(q)}$ with $|R_{i(q)} - R_{0(q)}| = s$ are tested.

The search result of HC using MCoff criterion ($\alpha = 1.7e^{-3}$) with jumping constraint on number of coefficients tested on BS-PA are shown in this section.

The merit values of each element in function of $Y$ (NMSE) and $R$ (Nb of Coeff) is depicted in Fig 3.25. The searching steps are also indicated in this figure. We can see that the searching path converges directly towards the low merit value region. The best structure is given as $X_{min}$ in this figure. The solution has its parameters which are

$$\begin{aligned} \mathcal{K}_a &= 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b &= 3, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c &= 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1 \end{aligned} \tag{3.28}$$

with 37 coefficients having NMSE of -32.65 dB. The number of coefficients jumping distance begins with $s = 2$ and ends with $s = 4$.

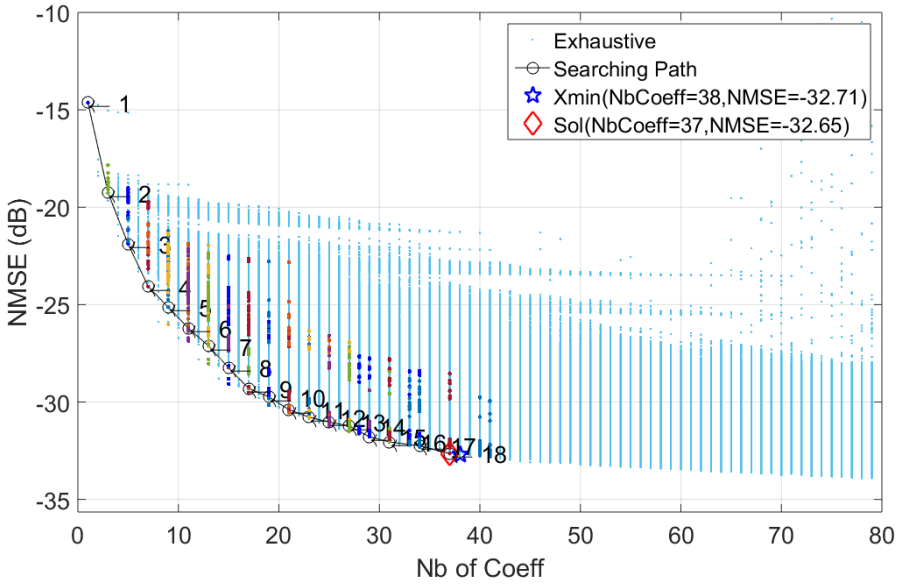



Figure 3.25: Seaching path of Hill-Climbing with jumping on Number of Coefficients

In this search, the algorithm has 18 iterations and there are totally 1438 different GMP model structures tested during 1.49 minutes. The merit values of each element in function of $Y$ (NMSE) and $R$ (Nb of Coeff) is depicted in Fig 3.25. The current elements of each iteration steps are indicated by circles with annotations in this figure. We can see that the searching path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$ in this figure.

In 2-step search, we first test 121 different MP models. The structure of $\mathcal{K}_a = 11$, $\mathcal{L}_a = 3$ with -31.13 dB NMSE is the best MP model which minimize (3.2). There are then 725 different GMP model structures tested in 7 iterations during 1.25 minutes. The searching path in map of NMSE vs number of coefficients is depicted in Fig 3.26. The blue points are again the exhaustive results, and the circles with annotations are the current element of each step in the Hill-Climbing

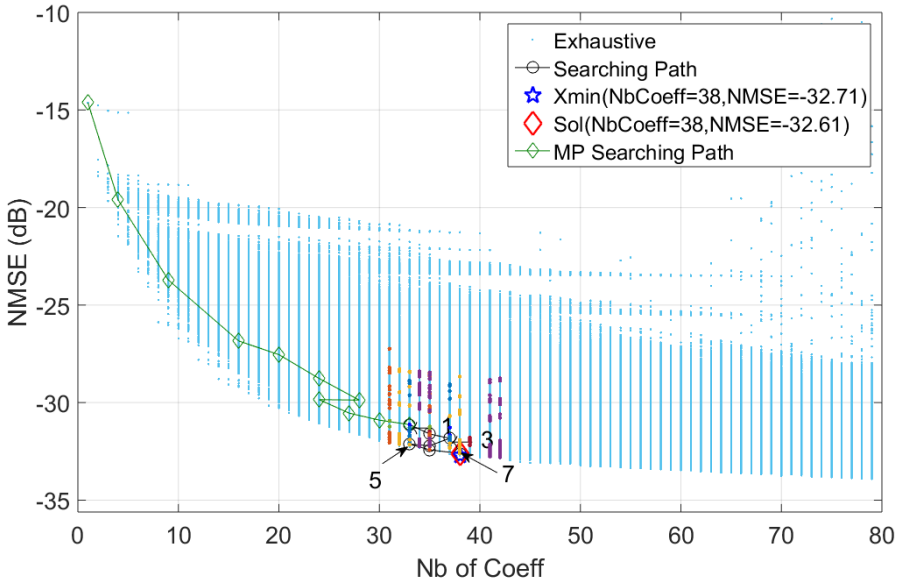



Figure 3.26: Two-step searching path of Hill-Climbing with jumping on Number of Coefficients

heuristic. The best structure found by the algorithm is

$$\begin{aligned} \mathcal{K}_a =& 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 2, \mathcal{M}_b = 2 \\ \mathcal{K}_c =& 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1 \end{aligned} \tag{3.29}$$

with 38 coefficients having NMSE of -32.61 dB. The number of coefficients jumping distance begins with $s = 2$ and ends with $s = 4$.

### HC with Unidimensional Neighbor

In this test, for the actual solution $x_{0(q)}$ at the $q$-th iteration, only the elements $x_{i(q)}$ with the variation of only one parameter are tested.

The search result of HC using MCoff criterion ($\alpha = 1.7e^{-3}$) with unidimensional neighbor tested on BS-PA are shown in this section.

The merit values of each element in function of $Y$ (NMSE) and $R$ (Nb of Coeff) is depicted in Fig 3.27. The searching steps are also indicated in this figure. We can see that the searching path converges directly towards the low merit value

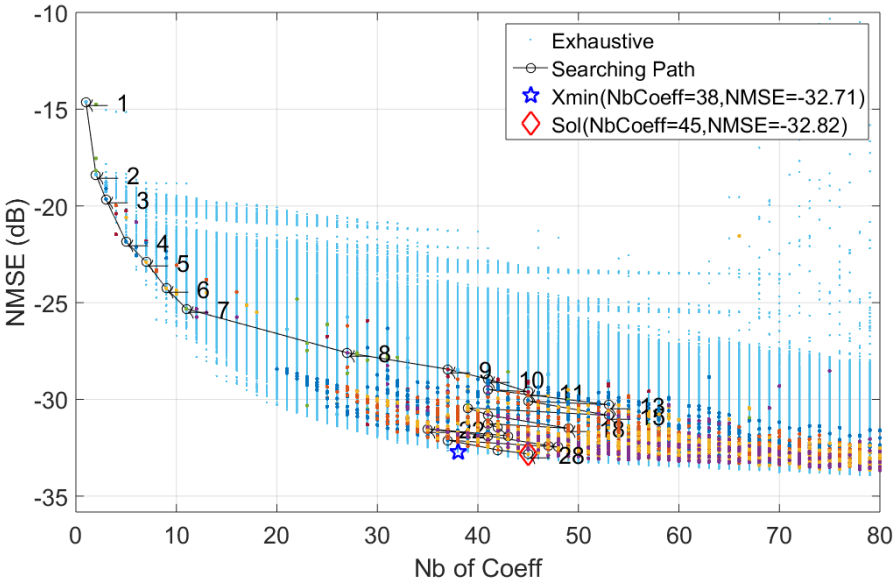



Figure 3.27: Seaching path of Hill-Climbing with Unidimensional neighborhood

region. The best structure is given as $X_{min}$ in this figure. The solution has its parameters which are

$$\begin{aligned} \mathcal{K}_a =& 2, \mathcal{L}_a = 4 \\ \mathcal{K}_b =& 10, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 9, \mathcal{L}_c = 3, \mathcal{M}_c = 1 \end{aligned} \tag{3.30}$$

with 45 coefficients having NMSE of -32.82 dB.

In this search, the algorithm has 10 iterations and there are totally 3575 different GMP model structures tested during 11.30 minutes. The merit values of each element in function of $Y$ (NMSE) and $R$ (Nb of Coeff) is depicted in Fig 3.27. The current elements of each iteration steps are indicated by circles with annotations in this figure. We can see that the searching path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$ in this figure.

In 2-step search, we can reach the best element in 12 iterations starting from MP model with $\mathcal{K}_a = 11$, $\mathcal{L}_a = 3$. There are 1393 different GMP model structures tested during 3.21 minutes. The searching path in map of NMSE vs number of coefficients is depicted in Fig 3.28. The solution is the same as (3.21) with 38

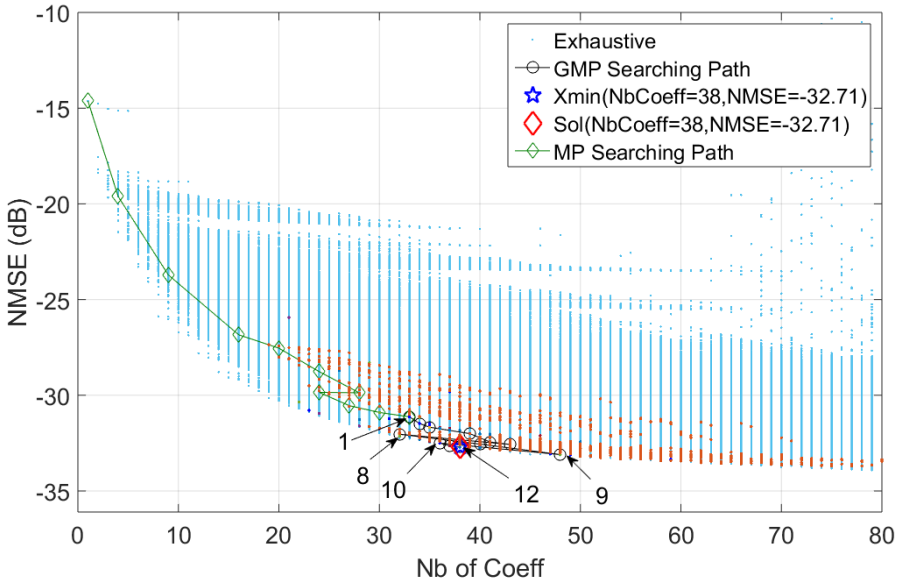



Figure 3.28: Two-step searching path of Hill-Climbing with Unidimensional neighborhood

coefficients having NMSE of -32.71 dB.

### Comparison of results obtained with pruned neighborhoods

The optimized algorithms are tested only on BS-PA with MCoff criterion. The execution time of conventional HC MCoff is taken as the reference to calculate the normalized execution times of the optimized algorithms in Table 3.15. By applying the three different pruned neighborhoods, the execution time is largely reduced. The result of constraint on number of coefficients shows that the best search path goes along the envelope. According to the result of unidimensional neighborhood, the search paths passing away from the envelope have heavy computational complexity even though the neighborhood size has been tremendously reduced. Jumping constraint on number of coefficients has the least execution time to converge. But as it jumps with a large step, the global optimum may be skipped.

The GMP models found by conventional HC with multiplicative criterion and HC with pruned neighborhoods are evaluated on test bench. Their spectra are illustrated in Fig 3.29. To evaluate the in-band and out-of-band linearization perfor-

Table 3.15: Comparison of Search Results of HC with Different Optimizations on BS-PA

Conv(_2): Conventional algorithm using MCoff criterion(two-step)
Optim1(_2): Constraint on NbCoeff is applied(two-step)
Optim2(_2): Jumping constraint on NbCoeff is applied(two-step)
Optim3(_2): Neighborhood is unidimensional(two-step)

| | Conv | Optim1 | Optim2 | Optim3 |
|---|---|---|---|---|
| $NMSE_{dB}$ | -32.71 | -32.71 | -32.65 | -32.82 |
| NbCoeff | 38 | 38 | 37 | 45 |
| Nb of Test | 19598 | 3835 | 1438 | 3575 |
| Nb of Iter | 11 | 21 | 18 | 10 |
| Ex. Time | 1 | 0.16 | 0.08 | 0.38 |
| | Conv_2 | Optim1_2 | Optim2_2 | Optim3_2 |
| $NMSE_{dB}$ | -33.04 | -32.71 | -32.61 | -32.71 |
| NbCoeff | 44 | 38 | 38 | 38 |
| Nb of Test | 39+7364 | 39+959 | 39+725 | 39+1393 |
| Nb of Iter | 11+4 | 11+7 | 11+7 | 11+12 |
| Ex. Time | 0.46 | 0.06 | 0.04 | 0.11 |

mances, error vector magnitude (EVM) and adjacent channel power ratio (ACPR) are listed in Table 3.16.

According to Fig 3.29 and Table 3.16, we can see that the linearization performances of the GMP models found by different means are close to each other. This confirms the efficiency of the proposed method in searching for the optimal GMP structures. And the proposed optimization methods do not deteriorate the robustness of the algorithm but tremendously improve the execution time.

All these optimizations are based on controlling the definition of the neighborhood during the search, which confirms the advantage of HC on the other algorithms with randomness.

Conv MCoff: Conventional HC with multiplicative criterion (3.21)
Conv MCoff2: Conventional HC with multiplicative criterion in 2-step (3.22)
Optim2: HC with jumping constraint on number of coefficients (3.28)
Optim2_2: HC with jumping constraint on number of coefficients in 2-step (3.29)
Optim3: HC with unidimensional neighborhood (3.30)

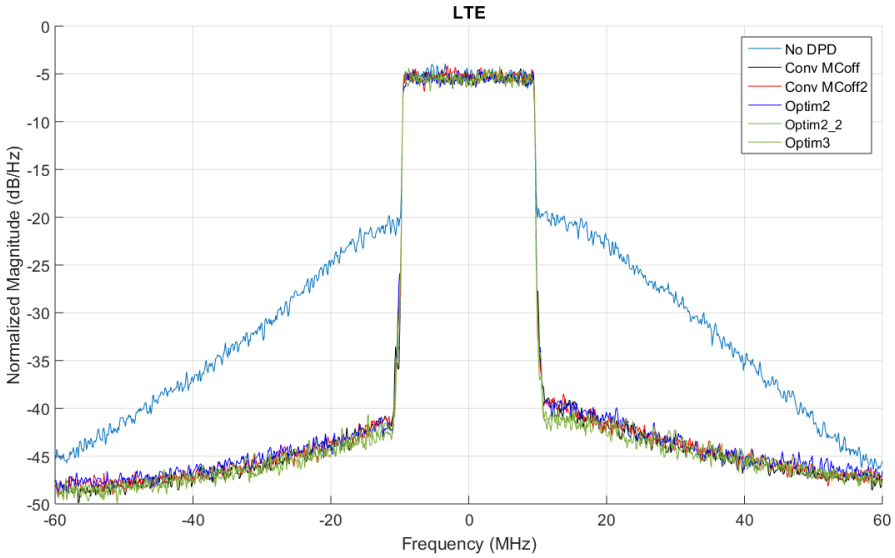



Figure 3.29: Comparison of spectra of Doherty PA output linearized by different DPD

Table 3.16: Performance comparison of GMP model solutions

| | GMP (3.21) | GMP (3.22) | GMP (3.28) | GMP (3.29) | GMP (3.30) |
|---|---|---|---|---|---|
| Nb of Coeff | 38 | 44 | 37 | 38 | 45 |
| EVM(%) | 3.4 | 4.8 | 3.3 | 4.0 | 3.0 |
| ACPR L1 gain(dB) | 18.4 | 18.3 | 17.8 | 18.9 | 18.5 |
| ACPR U1 gain(dB) | 18.1 | 17.8 | 17.8 | 18.9 | 18.2 |

## 3.8 Comparison Between Genetic Algorithm and Hill-Climbing Heuristic

### 3.8.1 Integer Genetic algorithm

A method using an integer genetic algorithm (GA) to select the orders of nonlinearities and memory lengths of the GMP model has been proposed in [62].

Genetic algorithms are stochastic search mechanisms inspired by the process of natural evolution. It is based on the idea of Darwinian natural selection theory that individuals with low values of quality will be eliminated and only those genes of good quality can be preserved in successive generations [78]. The quality is modeled by a fitness function in GA. In our implementation, the fitness function is given by the additive criterion (3.1).

To size the GMP model, the standard GA with functions for generating integer population and integer mutations is used in [62]. A vector of integer numbers represents the structure of a GMP model. The applied GA algorithm is described in Algorithm 3.

The initial population **P** individuals is generated as a random vector of integer numbers ($\mathcal{K}_a$, $\mathcal{K}_b$, $\mathcal{K}_c$, $\mathcal{L}_a$, $\mathcal{L}_b$, $\mathcal{L}_c$, $\mathcal{M}_b$, $\mathcal{M}_c$) which represents a chromosome made of genes at the beginning of the algorithm. The fitness function of the individual $x_{i(q)}$ at the $q$-th generation is denoted by $J_{i(q)}$. And $Q$ is the maximum generation number that the algorithm can reach.

At each generation, the offspring is created in three ways: selection, crossover and mutation. The groups of created offspring are denoted by $\mathbf{O}_1$, $\mathbf{O}_2$ and $\mathbf{O}_3$ respectively.

- Selection: The best 5% of **P** are kept in the next generation as offspring $\mathbf{O}_1$.
- Crossover: The parents can be represented in binary. For the parent $P_i$, its binary vector $\boldsymbol{v}_i$ are subdividing into groups $\boldsymbol{v}_i = [\boldsymbol{g}_{1,i}\ \boldsymbol{g}_{2,i}]$, and same for other parents. The crossover between $P_i$ and $P_j$ is to regroup $\boldsymbol{g}_{1,i}$, $\boldsymbol{g}_{2,i}$ and $\boldsymbol{g}_{1,j}$, $\boldsymbol{g}_{1,j}$ into new vectors, e.g. $[\boldsymbol{g}_{1,i}\ \boldsymbol{g}_{2,j}]$ and $[\boldsymbol{g}_{2,j}\ \boldsymbol{g}_{1,j}]$, etc.
- Mutation: The binary vectors can have some random variation on the bits.

The parents of the following generation is the sum of $\mathbf{O}_1$, $\mathbf{O}_2$ and $\mathbf{O}_3$.

The Stall condition is that the number of iterations reaches the limit or the fitness values variance of the population is lower than a threshold.

In this section, we compare the performances of these two methods for the sizing of GMP model structure using the same criterion based on trade-off between its modeling accuracy and complexity.

**Algorithm 3:** Overview of Genetic Algorithm

```
Set generation counter q = 1;
Generate initial random integer populations of P individuals;
while q ⩽ Q do
    Compute fitness of all individuals;
    Selection: The best 5% of P are kept as offspring O_1;
    Crossover: The 80% of the rest (P − O_1) are subdivided into groups
      and combined with each other to create offspring O_2 ;
    Mutation: for the rest, their genes in a chromosome are randomly
      changed to create offspring O_3;
    q=q+1;
    P = O_1 + O_2 + O_3;
    if Stall then
        end while loop;
    end
end
```

### 3.8.2 Performance comparison

In this section, we test only the additive criterion. If we choose $\mu = 0.065$, the best GMP model according to (3.1) has 38 coefficients ($\mathcal{K}_a = 11, \mathcal{L}_a = 2, \mathcal{K}_b = 4, \mathcal{L}_b = 1, \mathcal{M}_b = 1, \mathcal{K}_c = 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1$) and -32.71 dB $\text{NMSE}_{\text{dB}}$.

The two algorithms (GA and HC) have pretty different behaviors. HC is a deterministic algorithm looking at each step for the neighbor which has a better merit function, it can therefore be sensitive to local extrema. On the other side the GA algorithm may be seen as a probabilistic search starting from random initial population randomly spread throughout the whole solution space. Thus it could be quite sensitive to the size of the population which in turn has an impact on the complexity.

The sensitivity of the GA to different population sizes $\mathbf{P}$ is illustrated in Table 3.17 while its sensitivity to the number of "generation" (iteration) is highlighted by Table 3.18. By analyzing Table 3.17, we can see that the population size has

Table 3.17: GMP Results of Genetic Algorithm with Different Generation Sizes ($\mu = 0.065$)

| Population Size | 50 | 100 | 150 | 200 |
|---|---|---|---|---|
| $NMSE_{dB}$ | -32.22 | -31.73 | -32.71 | -32.65 |
| NbCoeff | 40 | 34 | 38 | 37 |
| Nb of Gen. | 10 | 10 | 10 | 10 |
| Ex. Time | 5.48 min | 10.99 min | 14.63 min | 20.35 min |

Table 3.18: GMP Results of Genetic Algorithm with Different Total Numbers of Generation ($\mu = 0.065$)

| Population Size | 100 | 100 | 100 | 100 |
|---|---|---|---|---|
| $NMSE_{dB}$ | -32.50 | -32.74 | -32.34 | -32.48 |
| NbCoeff | 37 | 42 | 36 | 35 |
| Nb of Gen. | 15 | 20 | 25 | 30 |
| Ex. Time | 13.72 min | 17.81 min | 19.80 min | 22.52 min |

a strong impact for the convergence towards the global optimum. We can check that the final solution is able to converge to the optimum when the population size is equal to or larger than 150 in this case. Table 3.18 shows that the increase of the number of generations does not help to converge towards the global optimum as the increase of the population size.

The results obtained with the HC algorithm and its convergence behavior are summarized in the two left columns of Table 3.19 and in Fig 3.30. From this picture we can see that HC behaves like a descent algorithm with iterations.

Table 3.19 allow to make some comparisons between both algorithms. First we can check that HC is able to converge to the global optimum while GA converges to a solution very close to this optimum solution but with a slightly higher complexity. The right column of Table 3.19 highlights that, unsurprisingly the two step approach does not improve the convergence speed and behavior of the GA algorithm, likely because of its randomness.

The lower subtable of Table 3.19 shows that different walks of the GA algorithms could lead to pretty different solutions. As the GA is a random search procedure, it may be improved by launching multiple times the GA search, picking the best solution among the results. This will make its performance more stable, but this will increase the execution time.

Table 3.19: GMP Results Comparison between Hill-Climbing Algorithm and Genetic Algorithm

| | HC | HC2 |
|---|---|---|
| Population Size | - | - |
| $NMSE_{dB}$ | -32.71 | -32.71 |
| NbCoeff | 38 | 38 |
| Structure | $\mathcal{K}_a = 11, \mathcal{L}_a = 2$<br>$\mathcal{K}_b = 4, \mathcal{L}_b = 1, \mathcal{M}_b = 1$<br>$\mathcal{K}_c = 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1$ | $\mathcal{K}_a = 11, \mathcal{L}_a = 2$<br>$\mathcal{K}_b = 4, \mathcal{L}_b = 1, \mathcal{M}_b = 1$<br>$\mathcal{K}_c = 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1$ |
| Nb of Iter/Gen | 11 | 6 |
| Ex. Time | 10.67 min | 7.74 min |
| | GA | GA2 |
| Population Size | 150 | 150 |
| $NMSE_{dB}$ | -32.88 | -32.48 |
| NbCoeff | 42 | 38 |
| Structure | $\mathcal{K}_a = 11, \mathcal{L}_a = 2$<br>$\mathcal{K}_b = 4, \mathcal{L}_b = 1, \mathcal{M}_b = 1$<br>$\mathcal{K}_c = 4, \mathcal{L}_c = 4, \mathcal{M}_c = 1$ | $\mathcal{K}_a = 10, \mathcal{L}_a = 2$<br>$\mathcal{K}_b = 1, \mathcal{L}_b = 3, \mathcal{M}_b = 2$<br>$\mathcal{K}_c = 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1$ |
| Nb of Iter/Gen | 10 | 10 |
| Ex. Time | 14.63 min | 16.36 min |

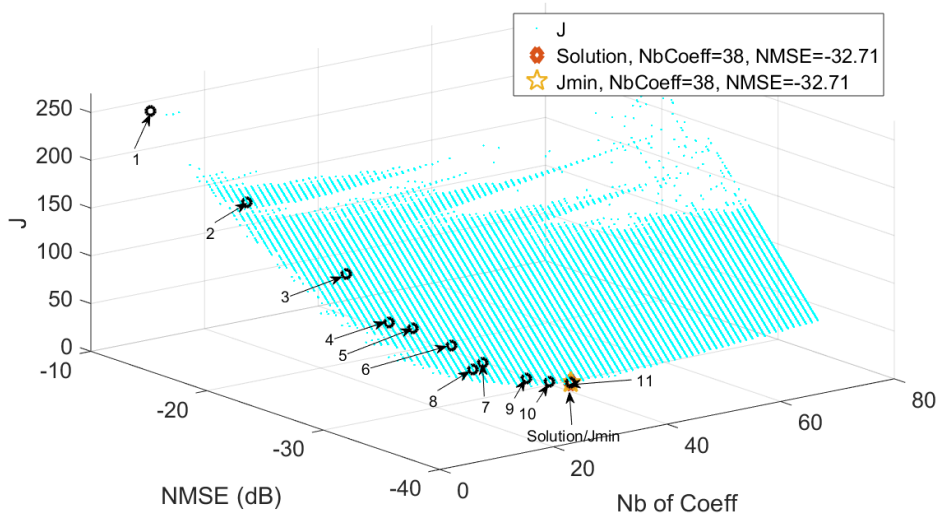



Figure 3.30: Hill-Climbing heuristic searching path in function of NMSE and number of coefficients in 3D ($\mu = 0.065$)

### 3.8.3 Conclusion

GA is an algorithm strongly depending on user settings, e.g. population size and total number of generation. Large population size helps approaching the best GMP result, but this also increases the execution time. With some appropriate settings, GA and HC have very approximate performances which converge to the global optimum in a short time compared with exhaustive search.

The advantage of HC as well demonstrated in the previous sections is that we can optimize the searching time by controlling the search path. But GA cannot be controlled because of its randomness.

## 3.9 First-Choice Hill Climbing

For the HC technique, we can also reduce the number of tested elements by controlling the search procedure among the neighbors, e.g. First-choice technique.

To reduce the number of tests during the search, we propose to test the neighbors under a certain condition, instead of testing all of them. The first tested neighbor which has better merit value than that of the current solution will be taken directly as the solution and the rest of neighborhood will not be tested. A

new iteration starts immediately. The neighbors are tested one by one following a random order, e.g. uniform law. The search procedure terminates when there is no neighbor better than the current solution. Then this solution is the solution of the algorithm. In this procedure, we added randomness to the algorithm and the final solution will not be always the same for each search.

This optimization method differs from the formers as it changes the search procedure instead of the neighborhood definition. However the induced randomness decreases its potential to be controlled and improved furthermore.

The light blue points in Fig 3.31 are all the possible GMP models in function of $Y$ ($\text{NMSE}_{\text{dB}}$) and $R$ (number of coefficients) in the search space. The colorful points represents the models tested during the search.

Using HC with ACoff, 10821 models are tested during 15.9 minutes and the solution is (3.21).

With FCHC, we have 16 iterations and totally 1691 structures of different GMP models are tested. The solutions of each iteration are indicated by the circles on the search path as shown in Fig 3.31. The search path converges towards the optimum $X_{min}$. The solution found by the algorithm is

$$\begin{aligned} \mathcal{K}_a =& 3, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 4, \mathcal{L}_b = 1, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 10, \mathcal{L}_c = 3, \mathcal{M}_c = 1 \end{aligned} \tag{3.31}$$

with 40 coefficients and $\text{NMSE}_{\text{dB}}$ of $-32.28$ dB. The execution time is 4.2 minutes.

As in this algorithm the neighbors are randomly chosen, We launch more times to verify the stability of the algorithm. Fig 3.32 shows another search path which also converges towards $X_{min}$. There are totally 4382 models tested in 15 iterations during 5.6 minutes. The final solution is

$$\begin{aligned} \mathcal{K}_a =& 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 4, \mathcal{M}_b = 1 \\ \mathcal{K}_c =& 3, \mathcal{L}_c = 3, \mathcal{M}_c = 1 \end{aligned} \tag{3.32}$$

with 35 coefficients and $\text{NMSE}_{\text{dB}}$ of $-32.48$ dB.

We can see that the found solutions have different structures but their linearization performances and execution times in search are very close.

The structure of GMP models (3.31) and (3.32) are evaluated on PA. The spectra of linearized PA output are illustrated in Fig 3.33. The blue curve is the spectrum corresponding to model 3.31 and the red curve is the spectrum corresponding

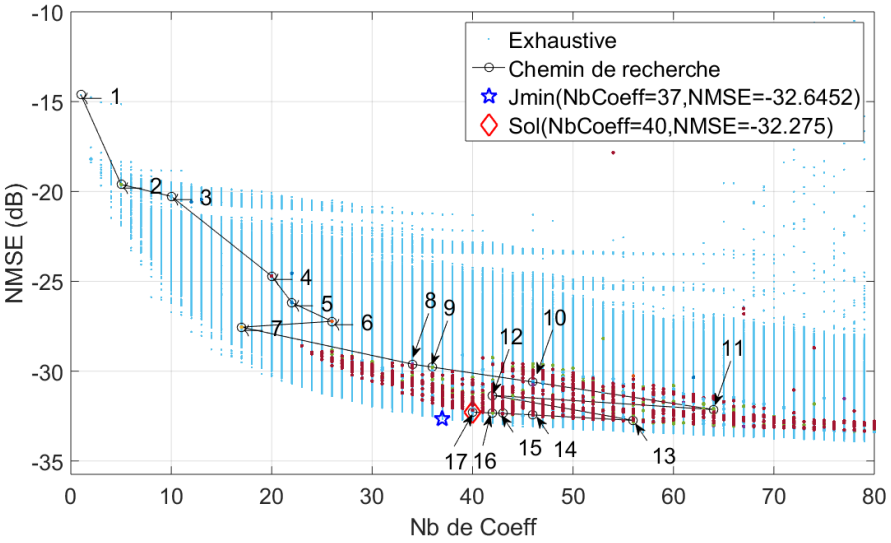



Figure 3.31: Result 1 of FCHC search

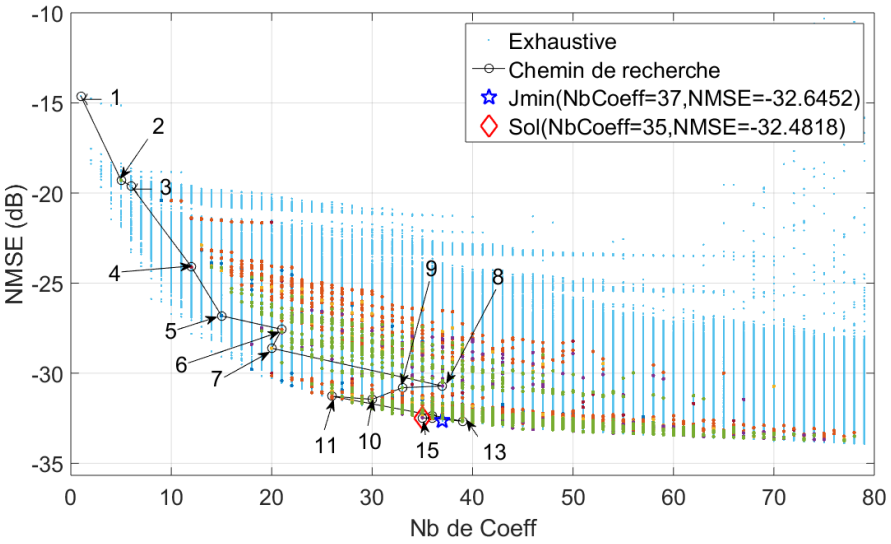



Figure 3.32: Result 2 of FCHC search

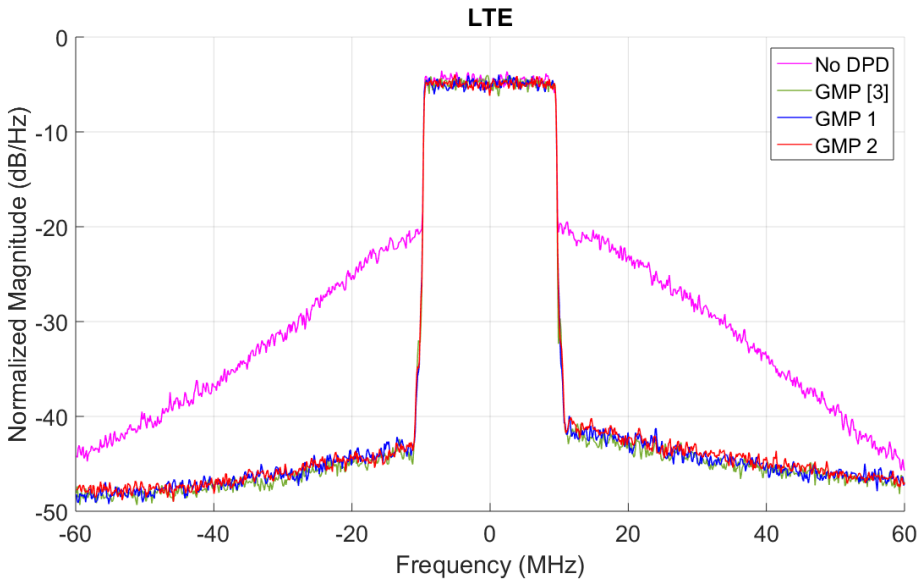



Figure 3.33: Comparison of spectra of Doherty PA output linearized by different DPD

to model 3.32. Their performances are very similar. The green curve is the spectrum of the reference GMP model (3.21).

To evaluate the in-band and out-of-band linearization performances, error vector magnitude (EVM) and adjacent channel power ratio (ACPR) are listed in Tab 3.20. The stability of FCHC is confirmed by the similarity of the performances of these two models.

We can see that the execution time of First-choice hill-climbing is only one third of that of conventional HC.

## 3.10 Conclusion

In this chapter, an algorithm is presented to find the optimal GMP model structure with a very limited number of computations. Two search criteria performing a trade-off between complexity and modeling accuracy of the GMP model are compared.

The algorithm is applied on two different dataset: signals of Base station PA and signals of broadcast PA. The characteristics of these two PAs and their input signals are very different. The performances of the proposed algorithm are good

Table 3.20: Performance comparison of GMP models obtained with FCHC

| | GMP (3.21) | GMP 1 | GMP 2 |
|---|---|---|---|
| Nb of Coeff | 38 | 40 | 35 |
| EVM(%) | 2.3 | 3.2 | 2.6 |
| ACPR L1 gain(dB) | 19.3 | 18.8 | 19.0 |
| ACPR U1 gain(dB) | 19.1 | 18.8 | 18.3 |
| Ex. Time | 15.9 min | 4.2 min | 5.6 min |

in both cases, which confirm the robustness of the algorithm.

Several optimizations of the algorithm are proposed and verified. The optimizations are made by giving different constraints on the neighborhood of search or on the search procedure. The results show that the computational complexity of search is dependent on the search path. When the search path follows the envelope edge of exhaustive result cloud, the complexity is lower compared with the case that search path is far from the envelope.

A comparison between hill-climbing and genetic algorithm is made. They may have similar performances when the parameters of algorithms are well configured.

The genetic algorithm relies more on the population size which has a great impact on convergence speed and computational complexity. One of the advantages of the hill-climbing algorithm compared with evolutionary algorithms is that the search path can be controlled by configuring the parameters, which results in different optimizations on execution time.

The technique to determine the optimal single-block DPD is proved robust in this Chapter and will be applied in case of cascaded DPD in the next chapter.

# Chapter 4

# Multi-stage Cascaded Digital Predistortion

## 4.1 Introduction

PD model can be not only one system, but also an union of several systems. Cascaded DPDs are found beneficial by some researches in [79], [3], [80], [81]. In [79], the cascaded model is identified in a nonparametric method. A parametric identification using multi-stage ILA is proposed in [3]. In [80], a two-stage DPD is used to compensate separately for high-order memoryless nonlinearities and low-order memory effects. The nonlinear bloc is realized with a look-up table (LUT) which is identified using ILA. The memory bloc is upstream of the nonlinear bloc, which is identified using DLA. In [81], a three-layer DPD is proposed as a bias information is introduced in the second layer.

Hammerstein, Wiener, and Wiener-Hammerstein models are widely used and their identification algorithms are thoroughly researched [82] [83]. In [82], the Wiener model is identified using particle swarm optimization (PSO) instead of least squares (LS). An enhanced Hammerstein model is proposed in [83] to suppress the spectrum regrowth caused by the memory effects. A branch parallel to the linear filter bloc is added where the input signal is multiplied by its magnitude. Each stage of the cascaded model is identified separately at different iterations.

Different multi-stage cascaded models have been studied in recent years, such as multi-stage Volterra Series [84], multi-stage MP models [85] or multi-stage GMP models [86].

The advantage of cascaded-structure PD is the relative simplicity of the low

order modules in terms of computational complexity and extensibility to the realization of higher-order functions [87]. Though the identification of cascaded DPD needs additional steps, the complexity increase of the identification task can be compensated for by the lower number of model coefficients [88].

This chapter is organized as follows. Block-oriented Nonlinear System identification methods are presented in Section 2. Identification methods of general multi-stage cascaded DPD are presented in Section 3. Section 4 gives some cascaded DPD performances experimental studies and analysis. An algorithm and criterion is used in Section 5 to find the optimal structure of cascaded DPD model. Section 6 gives a brief conclusion.

## 4.2 Block-oriented Nonlinear System Identification

Identification of multi-stage cascaded DPD is more complicated than that of single-stage model. A particular case of multi-stage cascaded DPD is the block-oriented nonlinear (BONL) system. The BONL structure is composed of blocks connected in series or in parallel: the Hammerstein model as Fig 2.9 and the Wiener model as Fig 2.10 are particular cases. The identification of Wiener or Hammerstein model has been studied in [89] [90] [91] [92].

The iterative identification algorithm of Hammerstein model proposed by Narendra and Gallman in [89] has been used in [93] for DPD. The Hammerstein model in Fig 2.9 which can be described by (2.21) and (2.22) is identified by an iterative method alternately adjusting the coefficients $a_k$ and $b_l$ by least square (LS) method. In this algorithm, a relationship among the coefficients of nonlinear part, memory effect part and passing signals is built. First of all, we need to initialize the values of $a_k$. By knowing the coefficients of nonlinear part, we can calculate the coefficients of memory effect part, and vice versa. In the following iterations, we can calculate the coefficients of both two parts iteratively till the algorithm converges.

A non-iterative method for Hammerstein identification was proposed in [90]. In this method, the values of $c_{k,l} = \sum_{k=1}^{K-1} \sum_{l=0}^{L-1} a_k b_l$ are estimated in one iteration with the input and output signals of the system. Here $a_0$ is assumed unity and other coefficients $a_k$ ($k > 0$) are normalized by $a_0$. For each memory depth $l$ ($l \in [0; L-1]$), we can calculate a corresponding value of $a_{k(l)}$ by:

$$a_{k(l)} = \frac{c_{k,l}}{c_{0,l}} \tag{4.1}$$

where $k \in [1; K-1]$. If we denote the vector of coefficients by $\boldsymbol{a} = [a_1, .., a_{K-1}]$, there are $L$ different possible coefficient vectors $\boldsymbol{a}_{(i)}$ ($i \in [0; L-1]$). Each of them is tested and the one which generates the least error between the expected signal and measured signal is taken as the best. Comparing with iterative method, this method can reduce the computational complexity by limiting the iteration number to only one, but the matrix size is largely increased (the array $c_{k,l}$ has $K \times L$ elements where $a_k$ and $b_l$ have only $K$ and $L$ elements respectively). Thus the non-iterative method has advantage only when the orders of model structure $K$ and $L$ are lower than a certain threshold.

A discrete-time Hammerstein model is identified in [94]. The linear dynamic stage can be identified by the correlation between the input and output signals. Then the nonlinear stage is identified using kernel regression estimates which was proposed independently by Nadaraya [95] and Watson [96]. With the same algorithm, the Wiener model is identified in a nonparametric method in [97].

Both iterative method and the correlation-based method have been two most studied method for years. However, the convergence of iteration method is confirmed in [91] and it is preferable because the correlation-based method with finite data lengths can not cover any particular realization of the input process over a finite time period. The convergence of iterative identification of Hammerstein model is also confirmed in [98].

The identification of a Hammerstein-Wiener model (as Fig 2.12a) using maximum likelihood (ML) estimation is discussed in [92]. The parameters vector of the system are denoted by $\boldsymbol{\theta} = [\boldsymbol{a}^T, \boldsymbol{b}^T, \boldsymbol{c}^T]^T$ where $\boldsymbol{a}$ and $\boldsymbol{c}$ are coefficient vectors of nonlinear blocs and $\boldsymbol{b}$ is the coefficient vector of the middle linear filter. An estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg max_{\boldsymbol{\theta}} L_{\boldsymbol{\theta}}(Y_N) \tag{4.2}$$

where $Y_N$ are the measured output signal of the system, $N$ is the number of realizations. According to Bayes' rule, the likelihood can be written as

$$L_{\boldsymbol{\theta}}(Y_N) = \sum_{t=1}^{N} \log p_{\boldsymbol{\theta}}(y_t|Y_{t-1}),\ p_{\boldsymbol{\theta}}(y_1|Y_0) = p_{\boldsymbol{\theta}}(y_1) \tag{4.3}$$

where $y_N$ are the estimated system output, $p_{\boldsymbol{\theta}}(y_1)$ denotes the joint density of $y_1$, and $p_{\boldsymbol{\theta}}(y_t|Y_{t-1})$ is the prediction density for the realization $Y_{t-1}$.

The initial values of the parameters are very important for the convergence of ML to the global optimum. The best linear approximation (BLA) [99] and

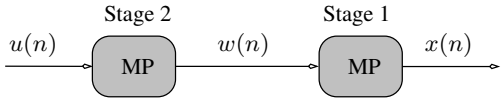



Figure 4.1: 2-stage MP model

$\epsilon$-approximation [100] can be used to extract the simple function of the linear dynamic bloc in the system. In [36], BLA is applied to identify different types of BONL with input signal $u(t)$ and output signal $y(t)$:

$$G_{bla}(e^{j\omega T_s}) = \frac{S_{YU}(e^{j\omega T_s})}{S_{UU}(e^{j\omega T_s})} \tag{4.4}$$

where $S_{YU}$ is the cross-power spectrum of $y(t)$ and $u(t)$, $S_{UU}$ is the auto-power spectrum of $u(t)$.

## 4.3 Identification of General Multi-stage DPD Model

For general multi-stage cascaded DPD models, we identify stage by stage as proposed in [3]. As multi-stage DPD is composed of several cascaded blocks, the block nearest to the PA is denoted by "Stage 1". We take a 2-stage MP model as an example in this dissertation (Fig 4.1).

The relation between the input and output of DPD can be expressed as:

$$\begin{aligned} w(n) &= \sum_{k=0}^{K_2-1} \sum_{l=0}^{L_2-1} e_{kl} u(n-l)|u(n-l)|^k \\ x(n) &= \sum_{k=0}^{K_1-1} \sum_{l=0}^{L_1-1} h_{kl} w(n-l)|w(n-l)|^k \end{aligned} \tag{4.5}$$

where $u(n)$ and $x(n)$ are the input and output signal of the DPD respectively, $w(n)$ is the output signal of Stage 2.

In case of a 2-stage MP as (4.5), we can model a DPD which has nonlinearities of order $K_1K_2$ and memory depth of order $L_1 + L_2 - 2$. The total number of coefficients is $K_1L_1 + K_2L_2$. But the values of the coefficients have very strong constraints. The coefficients of the equivalent predistorter structure comes from the product combinations of the coefficients of the 2 stages MP models.

The identification procedure using indirect learning structure (ILA) of cascaded DPD is presented in this section (Fig 4.2). A Post-inverse of each stage is

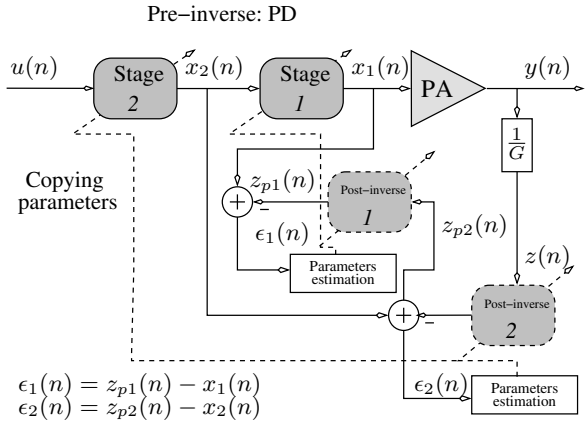



Figure 4.2: Indirect Learning Architecture - ILA

identified and used in the Pre-inverse block as a DPD. The identifications of both stages have the same procedure. Thus we take Stage 2 as an example.

The total number of coefficients of Stage 2 is $R_2 = K_2 L_2$. For a total number of samples equal to $N$, we can write

$$\boldsymbol{z_{p2}} = \boldsymbol{Z}\boldsymbol{a} \tag{4.6}$$

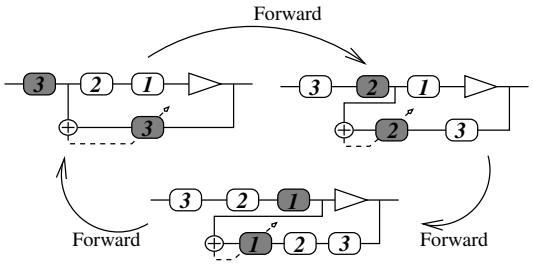
where $\boldsymbol{z_{p2}} = [z_{p2}(1), \ldots, z_{p2}(N)]^T$ is the vector of output samples of the post-inverse, $\mathbf{Z}$ is a $N \times R$ matrix of basis functions of $\boldsymbol{z}$, $\boldsymbol{a}$ is the vector of coefficients $a_{kl}$. The Least square (LS) solution of (4.6) is also the solution of the normal equation

$$[\mathbf{Z}^H \mathbf{Z}]\hat{\boldsymbol{a}} = \mathbf{Z}^H \boldsymbol{x_2} \tag{4.7}$$

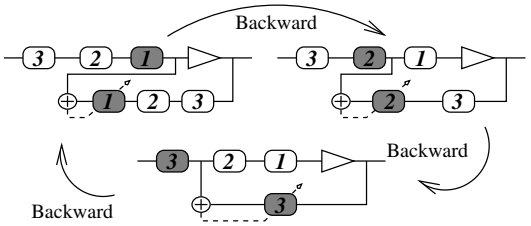
where $\hat{\boldsymbol{a}}$ is the estimated value of $\boldsymbol{a}$.

For the identification of multi-stage DPD, all stages can be identified with ILA iteratively in two ways: the order of identification from the stage furthest from the PA to the stage closest to the PA is called "forward (FW)"; (Fig 4.3a) the order of identification from the stage closest to the PA to the stage furthest from the PA is called "backward (BW)". (Fig 4.3b) The DPD stages are identified in an alternate way: after the last stage is identified, the following system iteration restarts from stage 1. We can re-identify the other stages after the modification of one stage.

In each system iteration, only one stage of DPD is identified while other stages are applied [3].




(a) Forward (FW) order




(b) Backward (BW) order

Figure 4.3: Identification order of multi-stage DPD

## 4.4 Experimental Results for Multi-stage MP model DPD

### 4.4.1 Case of a cascade of 2 low order MP models

In the following experiments, each stage is identified with 4 iterations. The DPD identification dataset buffer size is 6 000 samples. The DPD performance is evaluated on BS-PA with an LTE signal of 240 000 samples.

The simplest cascaded DPD is composed of only two stages. We choose the following cascaded MP structures:
Model K6L2/K2L6

$$\begin{aligned} K_1 &= 6, L_1 = 2 \\ K_2 &= 2, L_2 = 6 \end{aligned} \tag{4.8}$$

and Model K2L6/K6L2

$$\begin{aligned} K_1 &= 2, L_1 = 6 \\ K_2 &= 6, L_2 = 2. \end{aligned} \tag{4.9}$$

Both of them have 24 coefficients.

The computational complexity depends on the number of samples $N$ and on the total number of coefficients of the model. The main computation load comes from finding the solution of (4.7). We consider solving (4.7) using a QR decomposition. If N is large compared to $R$, we can estimate the computation load by $\mathcal{O}(2NR^2)$. As $2N$ is a constant, we express the complexity with $R^2$. The complexities of (4.9) and (4.8) are both 288.

In purpose of reference, we choose three MP models of similar complexity to test:
Model K6L4

$$K = 6, L = 4; \tag{4.10}$$

which has 24 coefficients with complexity equal to 576.
Model K6L3

$$K = 6, L = 3; \tag{4.11}$$

which has 18 coefficients with complexity equal to 324.
Model K5L3

$$K = 5, L = 3. \tag{4.12}$$

which has 15 coefficients with complexity equal to 225.

The model K6L4 is chosen because it has the same number of coefficients with the proposed cascaded MP model. The model K6L3 is chosen because it is similar to one stage of the proposed cascaded MP model and it has also a very close complexity to the proposed cascaded MP model. The model K5L3 is chosen because a model with complexity less than the proposed cascaded MP model needs to be tested as a reference.

The spectra of output signal of PA linearized with MP DPD (4.10)-(4.12) are shown in Fig 4.4. We can see that Model K6L4 has better linearization performance and it is taken as a reference in Fig 4.7 to compare with the cascaded MP models.

The spectra of output signal of PA linearized with Model K6L2/K2L6 in different identification orders are shown in Fig 4.5. The light blue curve is the output of PA without DPD. The linearization performance of BW is better than that of FW. The spectra of output signal of PA linearized with Model K2L6/K6L2 in different identification orders are shown in Fig 4.6. In this case, the linearization performance of FW is better than that of BW.

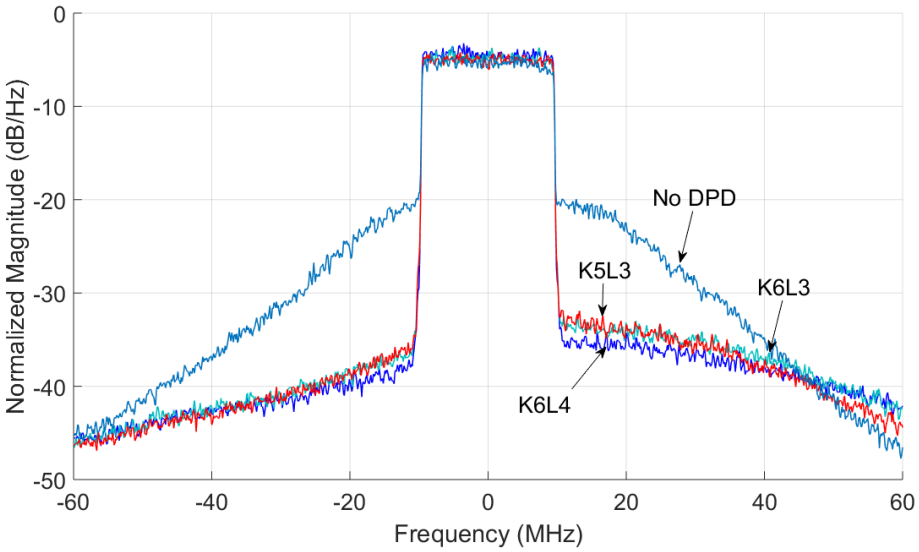


Figure 4.4: Spectra of Doherty PA output linearized by different MP DPD

In Fig 4.7, we compare the best results that the MP model and two cascaded models can achieve. We can see that Model K6L2/K2L6 and Model K2L6/K6L2 have the same performance and both of them outperform Model K6L4. This is confirmed in Table 4.1. In this table, the error vector magnitude (EVM) of the output signal and the first-adjacent-channel adjacent channel power ratio (ACPR) gain of the DPD identified with different algorithms are used to evaluate the linearization performances. The ACPR gain is the improvement of ACPR after linearization compared with the the output of PA without DPD.

Each stage of Model K6L2/K2L6 or Model K2L6/K6L2 have the same dimension, but one stage represents mainly nonlinearity and the other one represents mainly memory effect. Fig 4.5 and Fig 4.6 show that the identification of DPD starting from the model with the higher nonlinearity can always achieve the best performance of a given model. Fig 4.7 shows that cascaded DPDs with different structures can reach the same performance, which confirms that linearization performance of cascaded DPD depends not on the order of the blocks in the cascade.

### Conclusion

In this section, we analyzed the interest of using a 2-stage cascaded DPD composed of two different low order MP models. By switching the order of the

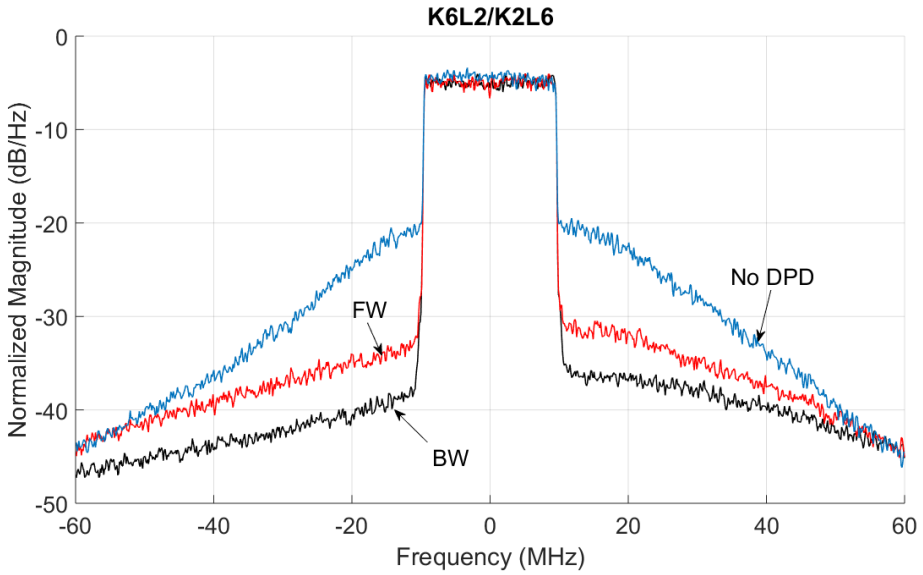



Figure 4.5: Spectra of Doherty PA output linearized by Model K6L2/K2L6 with different identification

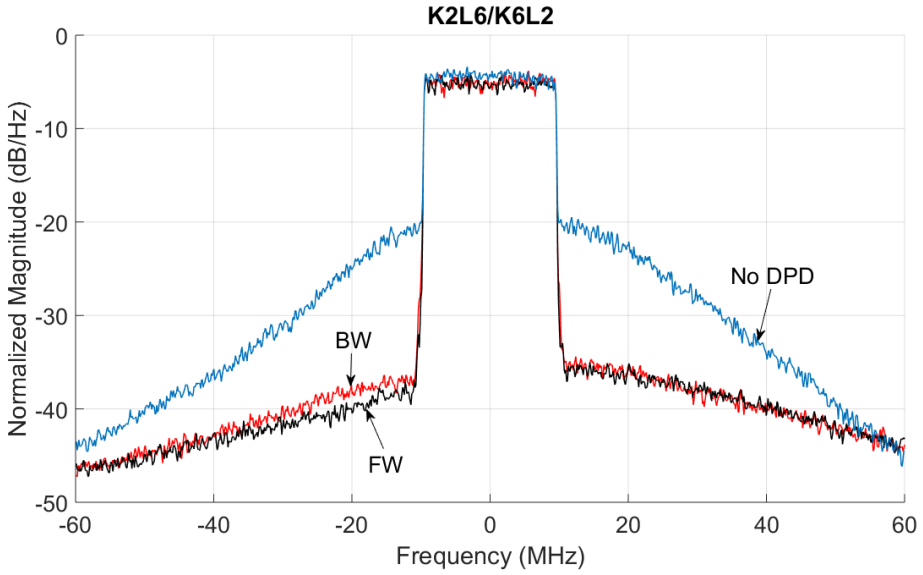



Figure 4.6: Spectra of Doherty PA output linearized by Model K2L6/K6L2 with different identification

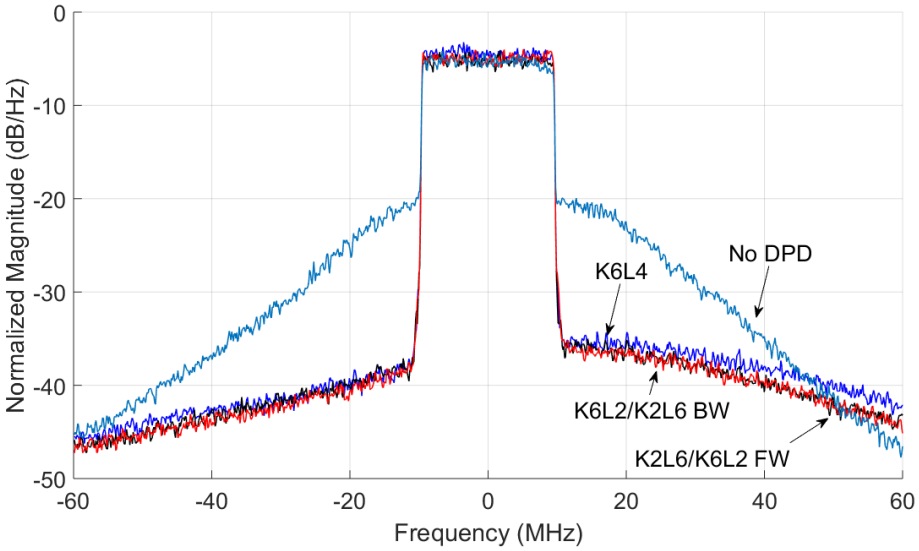



Figure 4.7: Comparison of spectra of Doherty PA output linearized by different DPD

Table 4.1: Performance comparison of 3 DPD models

| | ACPR L1 gain (dB) | ACPR U1 gain (dB) | EVM (%) | $J^2$ |
|---|---|---|---|---|
| K6L4 | 14.3 | 11.8 | 5.2 | 576 |
| K6L3 | 13.2 | 10.1 | 6.8 | 324 |
| K5L3 | 13.0 | 10.1 | 6.7 | 225 |
| K6L2/K2L6 FW | 9.7 | 8.7 | 9.7 | 288 |
| K6L2/K2L6 BW | 14.9 | 12.9 | 5.8 | 288 |
| K2L6/K6L2 FW | 14.2 | 12.4 | 5.3 | 288 |
| K2L6/K6L2 BW | 12.6 | 12.1 | 8.3 | 288 |

two stages, we compared their performances. We showed that the identification method of 2-stage cascaded DPD is of importance and the stage with higher nonlinearity order should be identified first. In the test case here, the switching of the stages does not alter the linearization performance provided that the right identification procedure is conducted.

### 4.4.2 Comparison of a multi-stage cascaded MP DPD with a single-stage MP or GMP model

This section makes an experimental study on comparison between a single-stage DPD and a multi-stage cascaded DPD model. The performance of cascaded DPD model is evaluated with the Three-way Doherty PA (BS-PA) and LTE signal.

To make a comparison between the single stage DPD and multi-stage DPD, we tested an MP model, a GMP model and a 2-stage MP model which have very similar number of coefficients.

Using our hill-climbing method, we can find an appropriate GMP model structure:

$$\begin{aligned} K_b =& 11, L_b = 2 \\ K_c =& 1, L_c = 3, M_c = 3 \\ K_d =& 3, L_d = 4, M_d = 1. \end{aligned} \tag{4.13}$$

As the GMP model has totally 43 coefficients, the MP model structure is set as

$$K = 11, L = 4. \tag{4.14}$$

The 2-stage MP structure is arbitrarily set as:

$$\begin{aligned} K_1 =& 9, L_1 = 4 \\ K_2 =& 4, L_2 = 2. \end{aligned} \tag{4.15}$$

In the following experiments, each stage is identified with 4 iterations. For example in case of BW, firstly we identify the stage 1 for 4 iterations. Then we fix the stage 1 and identify the stage 2 for 4 iterations. This procedure is denoted by "system iteration". In this section, the cascaded DPDs are identified with 2 system iterations.

The results of 8th iteration are shown in Fig 4.8 and Table 4.3. The DPD identification dataset buffer size is 6 000 samples. The DPD performance is evaluated with an LTE signal of 240 000 samples.

Table 4.2: Performance comparison of 3 DPD models

| | GMP | MP | 2-stage MP |
|---|---|---|---|
| EVM(%) | 2.34 | 5.86 | 3.66 |
| ACPR L1 gain(dB) | 19.1 | 16.1 | 17.1 |
| ACPR U1 gain(dB) | 18.1 | 15.3 | 16.0 |

The spectra of output signal of linearized PA are shown in Fig 4.8. The light blue curve is the output of PA without DPD. The red, green and black curves are the output of PA with DPD of 2-stage MP model, MP model and GMP model respectively.

These three DPDs have very close linearization performances. The GMP model DPD can reach the best linearization performance. Comparing with MP model, the performance of 2-stage MP is a little better.

Table 4.2 shows the EVM of the output signal and the first-adjacent-channel ACPR gain of the DPD identified with different algorithms compared with the output of PA without DPD while applying each model. We can see that 2-stage MP has the best in-band performance and the out-band performance is between the MP model and GMP model.

Table 4.3 shows the number of coefficients of each model, the dynamic range of model coefficients (the difference between the maximum and the minimum), the condition number of the matrix computation in identification procedure and the normalized complexity (Number of complex multiplications normalized by that of the MP case). The dynamic ranges of coefficients of GMP and MP models are large and their matrices exhibit large condition numbers. For "2-stage MP" model, the number of coefficients, EVM and ACPR are close to GMP and MP models, but its computational complexity is much less. As "2-stage MP" model has less order in nonlinearity and memory depth in each stage, its condition number is smaller by a factor of 100. The complexity of "2-stage MP" model identification is also reduced by more than 30%.

In this section, we analyzed the interest of using a multi-stage cascaded digital predistorter composed of two MP models. We compared its performance with that of a single stage predistorter made of an MP or GMP model. The optimal structure (orders of non-linearity and memory lengths) of the single stage DPD (MP or GMP) are determined using an HC, but the 2-stage MP model is determined in an ad-hoc way. GMP model reaches the best linearization performance but using a

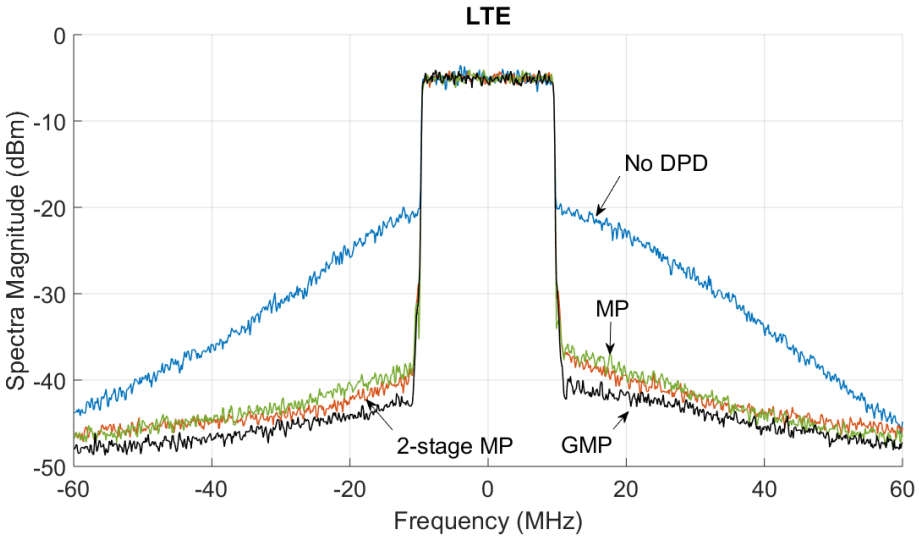



Figure 4.8: Comparison of spectra of Doherty PA output linearized by different DPD

Table 4.3: Performance comparison of 3 DPD models

| | GMP | MP | 2-stage MP | |
|---|---|---|---|---|
| | | | Stage 1 | Stage 2 |
| Number of Coefficients | 43 | 44 | 36 | 8 |
| Coeff Dynamic Range | $4.0e^6$ | $4.7e^6$ | $4.2e^5$ | 7.4 |
| Condition Number | $8.8e^{18}$ | $1.1e^{19}$ | $7.6e^{16}$ | $3.0e^6$ |
| Normalized Complexity | 0.96 | 1 | 0.67 | 0.04 |

2-stage MP model can decrease the complexity and dynamic range of coefficients and improve the condition number of the identification of the DPD parameters.

In this comparison, we have chosen arbitrarily the structure of the 2-stage DPD. In the next section, we propose a method to size the cascaded DPD structure.

## 4.5 Sizing of Cascade DPD Structure

In this section we apply the method proposed in Chapter 3 to size a cascaded model. We propose a new criterion to represent the trade-off between modeling accuracy and the complexity of model identification. A 2-stage MP model is chosen to be the cascaded DPD in this section. Its performance is evaluated with a Three-way Doherty PA.

### 4.5.1 Search Algorithm and Criterion

In a discrete set, each node $x_i$ is assigned to a unique cascaded MP model structure. The coordinate of $x_i$ consists of 4 dimensions: $K_1^{(i)}$, $L_1^{(i)}$, $K_2^{(i)}$, $L_2^{(i)}$. The value of a merit function $J(x_i)$ is associated to each node $x_i$.

The searching procedure is described in Algorithm 1.

In our implementation, we define a neighbor of node $x_i$ as a node of which parameters are $K_1^{(i)} + \delta_1$, $L_1^{(i)} + \delta_2$, $K_2^{(i)} + \delta_3$, $L_2^{(i)} + \delta_4$, where $\delta_{1,\dots,4} \in [0, \pm 1]$ and $\delta_{1,\dots,4}$ are not $0$ at the same time.

The merit value of the node $x_i$ here refers to a criterion $J_i$ which leads to a tradeoff between modeling accuracy and model complexity. NMSE value (2.7) denoted by $Y$, is the criterion to judge the modeling accuracy. The complexity of model identification is represented by the square sum of number of coefficients of the model which is denoted by $R_1^2 + R_2^2$.

We combine the objective on the number of coefficient and the constraint on the NMSE in a single criterion, using a trade-off weight coefficient $\lambda$:

$$J_i = Y_i + \lambda(R_{1,i}^2 + R_{2,i}^2). \tag{4.16}$$

### 4.5.2 Experimental Validation

In this section we evaluate our algorithm with 2 values of $\lambda$ corresponding to two trade-offs between the modeling accuracy and complexity of models identification. The values used for $\lambda$ are: $1e^{-3}$ and $2e^{-3}$.

In each case the algorithm starts from the node which corresponds to two linear functions $x_1(n) = bx_2(n)$ and $x_2(n) = au(n)$, where $K_1 = 1$, $L_1 = 0$, $K_2 = 1$, $L_2 = 0$.

With $\lambda = 1e^{-3}$, the algorithm has 11 iterations and there are totally 417 different cascaded MP model structures tested during 37.7 seconds. The merit values of each node in function of $Y$ (NMSE) and $R_1^2 + R_2^2$ is depicted in Fig 4.9. The current nodes of each iteration steps are indicated by circles in this figure. The light blue points are all possible cascaded MP models. The tested models are illustrated in colors. We can see that the searching path converges towards the lowest merit value which corresponds to the best structure given as $X_{min}$ in this figure. According to all these nodes, we found a structure having NMSE of $-31.13$ dB with the following parameters

$$\begin{aligned} K_1 = 11, L_1 = 3 \\ K_2 = 1, L_2 = 1. \end{aligned} \tag{4.17}$$

This structure is a particular case of cascaded model. It is equal to a 1-stage MP model because Stage 2 is only a gain.

With $\lambda = 2e^{-3}$, the first simulation also starts from the node which corresponds to simple gains.

In this search, the algorithm has 10 iterations and there are totally 368 different cascaded MP model structures tested during 35.6 seconds. The merit values of each node in function of $Y$ (NMSE) and $R_1^2 + R_2^2$ is depicted in Fig 4.10. We can see that the searching path converges towards $X_{min}$. According to all these nodes, we found a structure having NMSE of $-31.03$ dB with the following parameters

$$\begin{aligned} K_1 = 10, L_1 = 2 \\ K_2 = 3, L_2 = 3. \end{aligned} \tag{4.18}$$

In this structure, the two stages are not balanced. Stage 1 has much higher nonlinearity order than Stage 2 and their memory depths are similar.

The linearization performances of these two cascaded models ((4.17) and (4.18)) on test bench are shown in Fig 4.11 and Table 4.4. Both PDs can linearize the PA with an improvement of adjacent channel power ratio (ACPR) up to 16 dB and the error vector magnitude (EVM) is limited to 4%. In Table 4.4, ACPR L1 gain represents the improvement of ACPR in the first lower adjacent channel, and ACPR U1 gain represents that in the first upper adjacent channel. The two DPD have very similar performances but Model K10L2/K3L3 has less complexity.

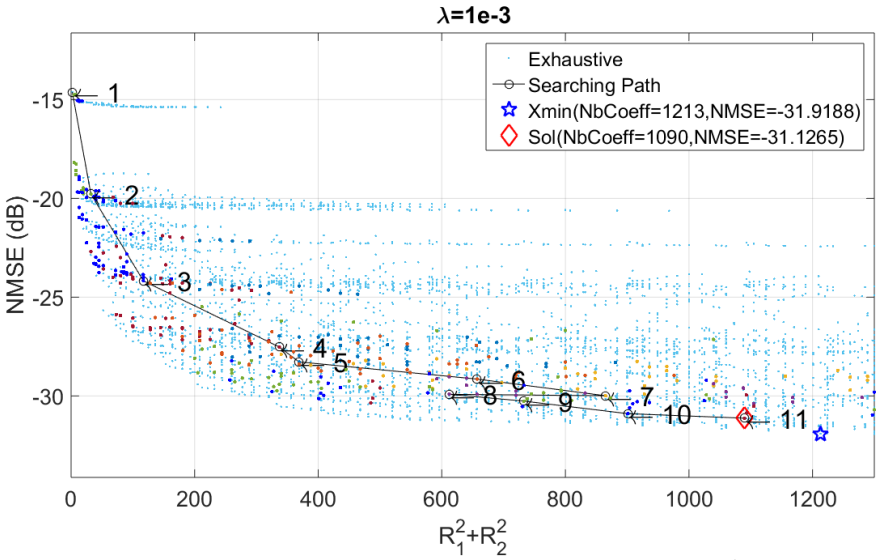



Figure 4.9: Cascaded model structure searching path with $\lambda$ equal to $1e^{-3}$

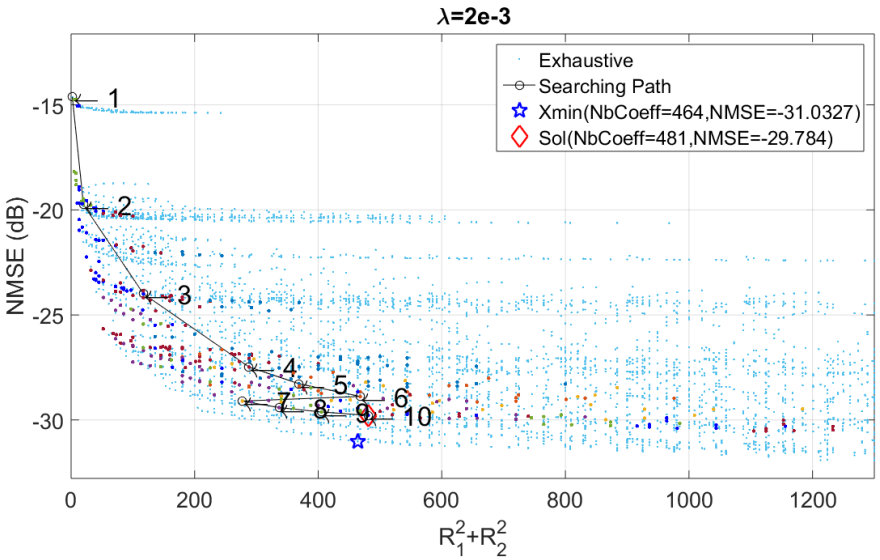



Figure 4.10: Cascaded model structure searching path with $\lambda$ equal to $2e^{-3}$

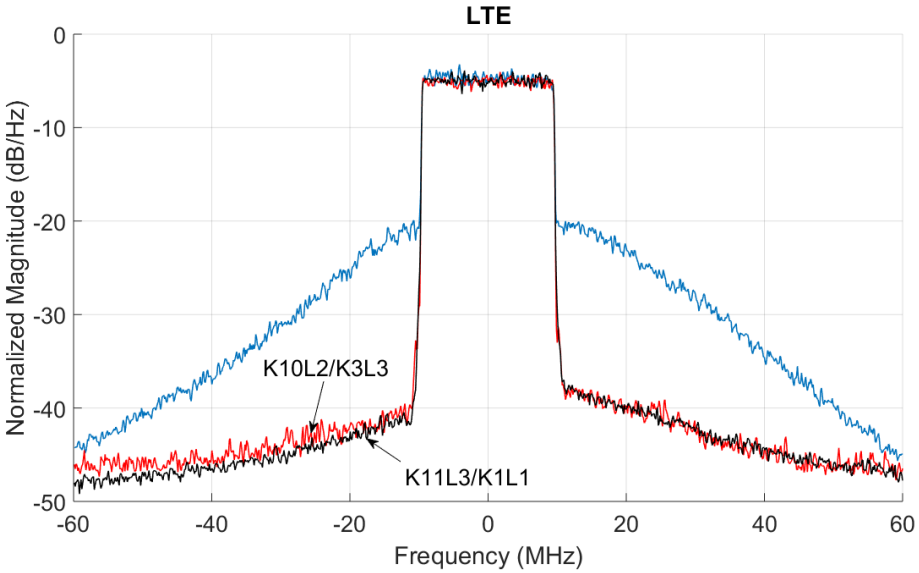



Figure 4.11: Comparison of spectra of Doherty PA output linearized by found DPD

Table 4.4: Performance comparison of 3 DPD models

| | ACPR L1 gain (dB) | ACPR U1 gain (dB) | EVM (%) | $R_1^2 + R_2^2$ |
|---|---|---|---|---|
| K11L3/K1L1 | 17.4 | 16.1 | 4.0 | 1090 |
| K10L2/K3L3 | 16.5 | 16.9 | 4.0 | 481 |

### 4.5.3 Conclusion

In this section we extend the search algorithm based on hill-climbing heuristic for a 2-stage MP models. We have also proposed a new criterion to search the structure of a 2-stage cascaded DPD composed of two MP models, which represents a trade-off between complexity of model identification and modeling accuracy. The solutions of the search algorithm are applied successfully to a three-way Doherty PA. Their linearization performances are robust.

## 4.6 Conclusion

This chapter thoroughly studied the linearization of PA using cascaded DPD. Experimental results show that the cascaded DPD may have advantages on computational complexity while keeping a good linearization performance compared with some single-block DPDs.

As the structure of cascaded DPD is more complicated, there are different ways to identify the coefficients of the model. Identifying stage by stage can estimate the coefficients by solving linear equations. Different identification orders can result in different performances. The DPD identified from the stage with the highest nonlinearity order has always the best performance.

The structure of cascaded DPD can be optimized according to a given criterion in a similar way to the algorithm presented in Chapter 3.

Next chapter explores the techniques of experimental implementation.

# Chapter 5

# Impact of PA Gain Choices

## 5.1 Introduction

The choice of the reference gain $G$ used in predistortion has been discussed in different papers [101] [102] [103] [104].

The linearized PA gain which impacts the DPD model can be decided arbitrarily [101]. In particular, two very often used gains are the small signal gain $G_1$ and the peak power gain $G_2$. The small signal gain $G_1$ is the gain of the PA when its behavior is generally linear. The peak power gain $G_2$ is calculated with the input peak power and output peak power of PA.

In [102], different normalization gains with look-up table (LUT) model DPD are tested. Zhu *et al.* [103] emphasize that using $G_1$ needs an extra calibration step and using $G_2$ does not. In [105], different normalization gains with memory polynomial (MP) model DPD are tested. The power alignment issue has been studied and the normalization gain for which the average powers at the input and output of the DPD is equal is considered optimal in [88]. In [106], $G_1$ is applied and the gain of DPD is set unity. A one-step method to extract the optimal normalization gain is proposed in [104]. The impact of choice of gains on the average output power of PA is studied in [107]. However, the scaling of gain is seldom discussed in terms of power added efficiency (PAE) in publications.

This chapter presents an experimental study of linearization and PAE with a Three-way Doherty PA at different choices of normalization gain. In Section 2, the power variation and PAE at the operating point are studied. And we propose an adjustment in the DPD identification to improve the PA linearization performance in case of $G_1$. The experimental results are presented in Section 3. Section 4 gives

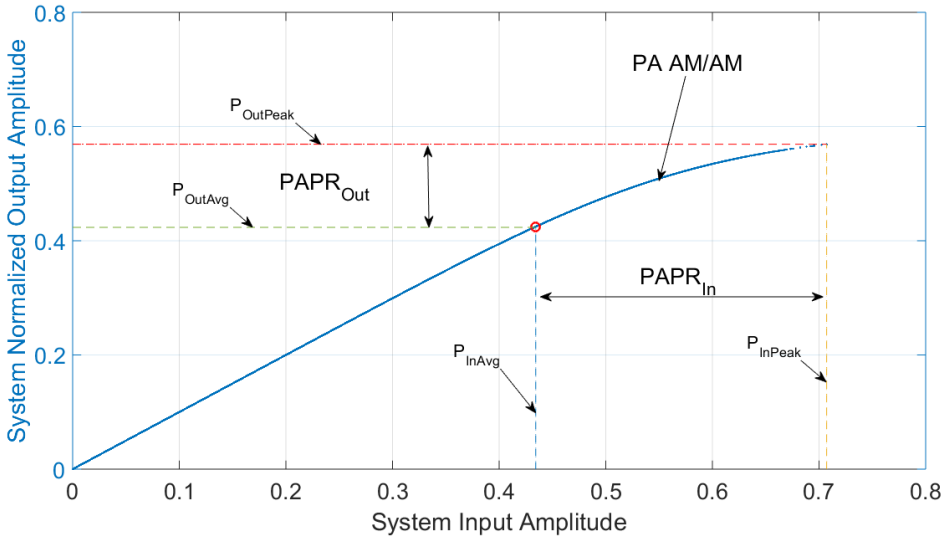



Figure 5.1: PA input signal vs PA output signal

a brief conclusion.

## 5.2 Linearization of PA at different gains

### 5.2.1 PA power compression effect

From Fig 2.14 we can see that: at the first iteration of identification, the PA input signal is $x(n) = u(n)$, and the post-inverse block is estimated with signal $z(n)$ which is the normalized output signal of PA; and at the second iteration, the estimated model of post-inverse block is applied at the pre-inverse block as the DPD while the DPD input signal is $u(n)$.

Because of the nonlinearity of PA near to the saturation zone as shown in Fig 5.1, the PAPR of PA output signal is compressed compared with the PAPR of PA input signal. We can easily have $PAPR_u > PAPR_z$. Thus the input signals of pre-inverse and post-inverse blocks are different.

As the signal $z(n)$ is normalized by the gain $G$ ($z(n) = y(n)/G$), the average power or peak power of $z(n)$ can be decided by $G$. If we respect the consistency

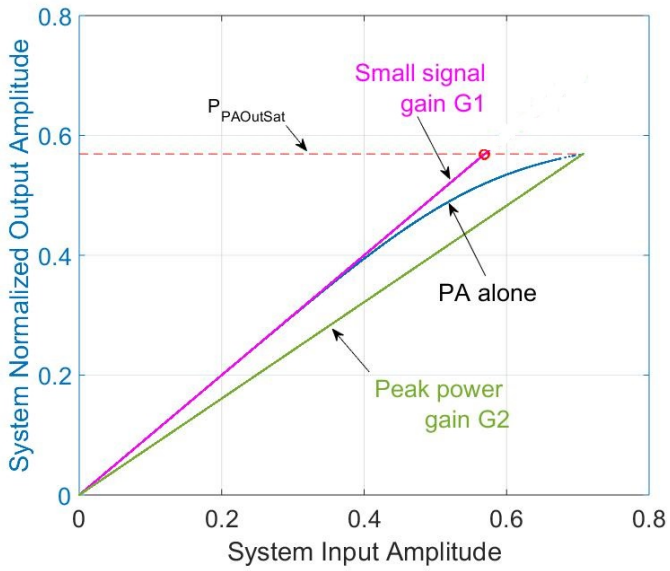



Figure 5.2: Two choices of gain

of average power, so we should have

$$E(P_u) = \frac{E(P_y)}{G}. \tag{5.1}$$

If we respect the consistency of peak power, so we should have

$$max(P_u) = \frac{max(P_y)}{G}. \tag{5.2}$$

Howerver, the consistencies of average power and peak power cannot be respected at the same time. We can also have different other choices of gain that none of them is respected.

In this dissertation, we choose only two gains as shown in Fig 5.2 to compare: small signal gain $G_1$ and peak power gain $G_2$. The output amplitude is normalized by the chosen gain $G_1$.

### 5.2.2 Linearization at $G_1$

There are many different ways to set the normalization gain which is used to calculate the coefficients of the DPD.

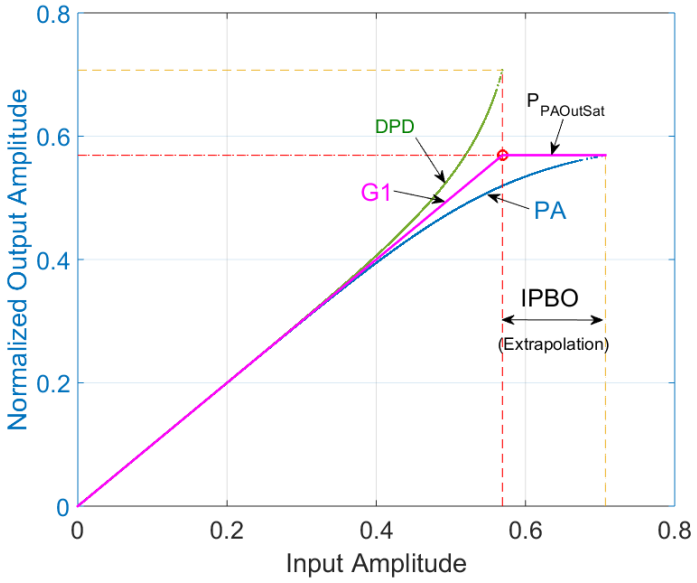



Figure 5.3: Small signal gain $G_1$

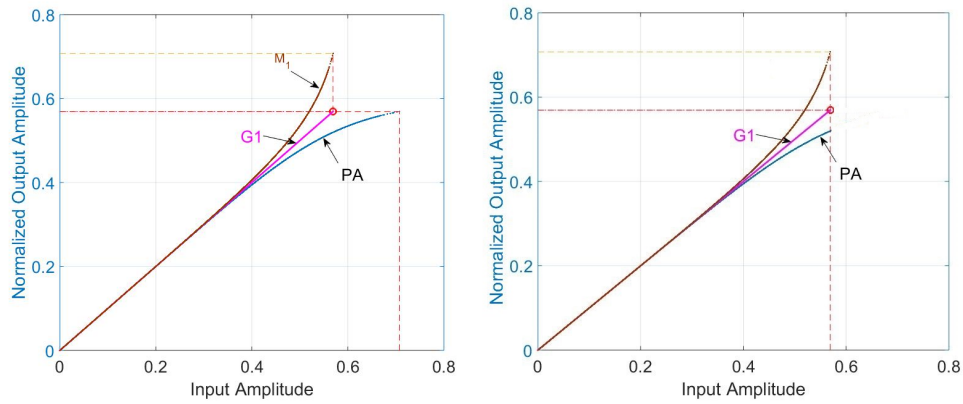
The small signal gain $G1$ is calculated using the low amplitude input signals. We can make an approximation $E(P_u) \approx E(P_z)$.

In Fig 5.3, the blue curve shows the AM/AM curve of PA, and the green curve is the AM/AM curve of DPD identified with input and output signals of PA. The pink line shows that the PA is linearized by the DPD. We can see that the output signal is saturated when the input signal amplitude is higher than a threshold. $P_{PAOutSat}$ is the saturated output of PA normalized by $G_1$.

Therefore, we need to apply an input peak back-off (IPBO) to reduce the dynamic range of the input signal. If we denote the threshold amplitude of the system input signal by $P_{th}$, a first step (calibration step) with a high power signal is needed to find out $P_{th}$. As $P_{PAOutSat}$ is the normalized PA saturated output amplitude, the input saturation power $P_{th}$ is obtained by

$$P_{th} = P_{PAOutSat}. \tag{5.3}$$

In this linearization method, the average power of the system input signal is kept at a constant value. In each iteration, the training signal $z(n)$ has peak power equals to $E(P_u) + PAPR_z$. In the following iterations, this DPD model is fed with a signal which has peak power equals to $P_{th} = E(P_u) + PAPR_u$. As $PAPR_u > PAPR_z$, the DPD model does not cover the entire power range of the input signal. An extrapolation of the DPD characteristic function is needed for




(a) Iteration 1: Identification without Back-off

(b) Iteration 2: Identification with Back-off

Figure 5.4: Adjustment for small signal gain $G_1$

the signals which have higher power than $E(P_u) + PAPR_z$. This degrades the performance of PA linearization.

### 5.2.3 Adjustment at $G_1$

As discussed in the previous subsection, when $G_1$ is chosen, the performance of the first identified DPD is degraded due to the extrapolation. The effect is evident when the PA has very strong nonlinearity. We propose an adjustment as shown in Fig 5.4:

- at the first iteration, a post-inverse model "$M_1$"is identified without IPBO;
- at the second iteration, IPBO is applied on the input signal and "$M_1$" is applied as DPD, and a new post-inverse model "$M_2$" is identified.

The peak power of training signal of "$M_1$" is the input peak power before IPBO. If we take this signal to train the post-inverse model, the input power range will be covered.

Then at the second iteration, the output signal of DPD will not diverge and the model "$M_2$" can be identified.

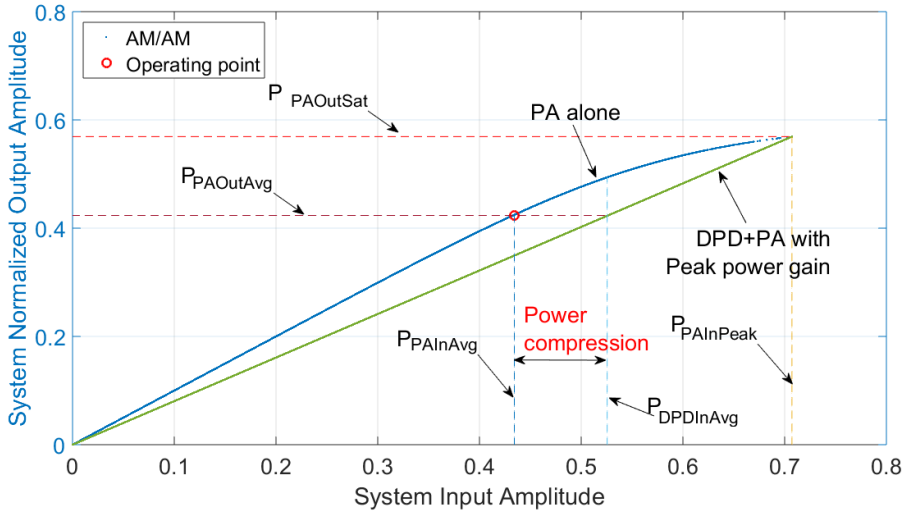



Figure 5.5: Peak power gain $G_2$

### 5.2.4 Linearization at $G_2$

In case of peak power gain $G_2$, the input power of PA should be high enough to push the PA to the critical point where it starts to be saturated. In Fig 5.5, the green straight line shows $G_2$ of the PA linearized by DPD which is calculated with the maximal amplitude of signals at input and output of PA:

$$G_{Peak} = \frac{P_y}{P_u}. \tag{5.4}$$

The advantage of this choice of gain is that no IPBO is needed and $u(n)$ and $z(n)$ have the same peak power. The drawback of this linearization method is that there is an average power difference between $u(n)$ and $z(n)$.

We can observe in Fig 5.5 that the average input power of DPD is higher than that of PA. In another word, the DPD compresses the power.

## 5.3 Experimental validation

The linearization performances are evaluated by the error vector magnitude (EVM) and the adjacent channel power ratio (ACPR) between the input signal of DPD and output signal of PA. The ACPR gain (ACPR G) is the improvement of ACPR of PA output after applying DPD compared with the ACPR of PA output without

DPD. The EVM is not the real EVM as (2.9). Since we process the signals in baseband, it is simply calculated by

$$\text{EVM}_{\%} = \frac{std(z(n) - u(n))}{std(u(n))} \times 100\% \tag{5.5}$$

where $u(n)$ is the stimulus signal, $z(n)$ is the normalized PA output signal, $std(\cdot)$ calculates the standard deviation.

Using the method proposed in Chapter 3, we can find an appropriate GMP model structure:

$$\begin{aligned} \mathcal{K}_a =& 11, \mathcal{L}_a = 2 \\ \mathcal{K}_b =& 1, \mathcal{L}_b = 3, \mathcal{M}_b = 3 \\ \mathcal{K}_c =& 3, \mathcal{L}_c = 4, \mathcal{M}_c = 1. \end{aligned} \tag{5.6}$$

### 5.3.1 Performance of linearization with adjustment at $G_1$

The measured spectra at the output of the linearized PA are presented in Fig 5.6. The green curve represents the spectrum of PA output signal without DPD. The blue curve shows the output signal spectrum of PA linearized at $G_1$. The red and the black curves are spectra of the output signal of linearized PA with the DPD identified during the calibration step and during the second step respectively. The spectra captured by FSW have 30 dB attenuation compared with the PA output power.

The measured ACPR gain and EVM are shown in Table 5.1. In the table, L1 refers to lower band 1st adjacent channel, U1 refers to upper band 1st adjacent channel, L2 refers to Lower band 2nd adjacent channel, U2 refers to Upper band 2nd adjacent channel.

We can see that the adjusted method improves the ACPR gain of the lower band 1st adjacent channel 9.52 dB, and the EVM is lowered from 15.3% to 6.2%.

Comparing these three spectra, we can see that the linearization with the adjustment outperforms the method without adjustment on the linearization of PA both in-band and out-of-band.

Table 5.1: Linearization performances of the PA at $G_1$

| | | Original Approach | Adjusted | |
|---|---|---|---|---|
| | | | 1st step | 2nd step |
| ACPR Gain(G) (dB) | L1 | 6.17 | 9.35 | 15.69 |
| | U1 | 4.99 | 12.84 | 15.69 |
| | L2 | 0.98 | 5.81 | 9.02 |
| | U2 | -2.95 | 4.61 | 6.75 |
| EVM (%) | | 13.3 | 23.2 | 4.2 |

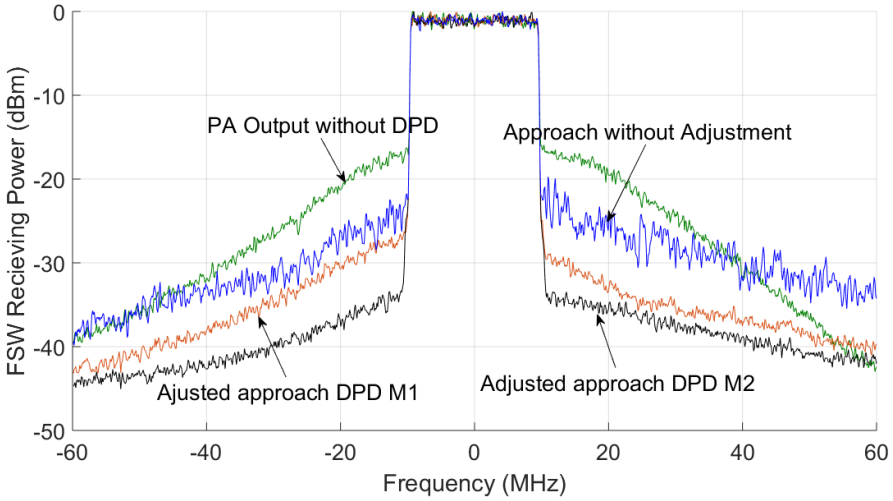



Figure 5.6: Measured spectra of the linearized PA

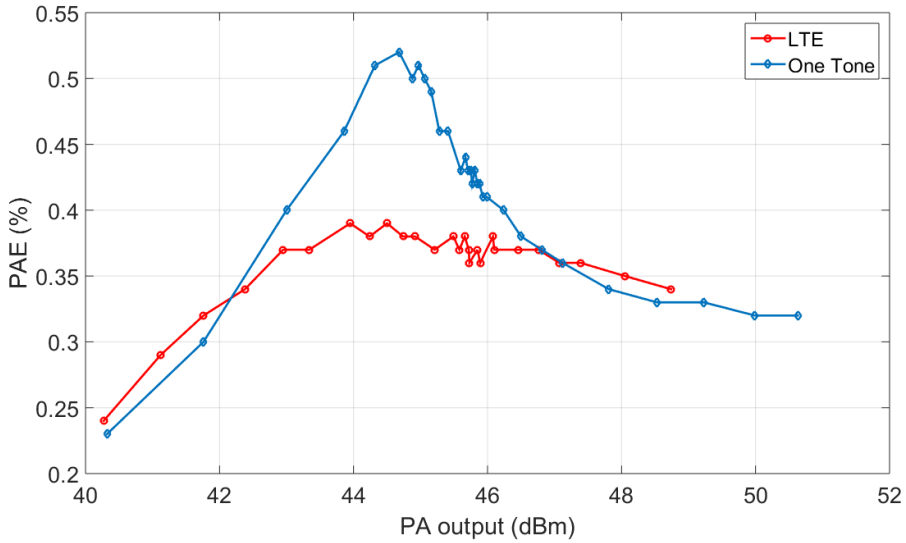



Figure 5.7: PAE of Doherty PA with different input signals

### 5.3.2 PAE vs PA output

The efficiency of the PA without DPD is measured as a reference. PAE is calculated by:

$$PAE = \frac{P_{out} - P_{in}}{P_{supply}}, \tag{5.7}$$

where $P_{in}$ and $P_{out}$ are system input and output power respectively, $P_{supply}$ is the power offered by supply.

In Fig 5.7, the blue and the red curve show the PAE tested with both one-tone signal and LTE signal respectively. The PAE keeps increasing when the average output power of PA is less than 44 dBm. When the average output power of PA is more than 47 dBm, we notice that PAE begins to decrease. To avoid the risk the overload of PA, we need to keep the peak power of PA output signal less than 57 dBm.

We chose 2 operating points to evaluate the performance of the 3 linearization methods: output at 45 dBm and output at 46 dBm.

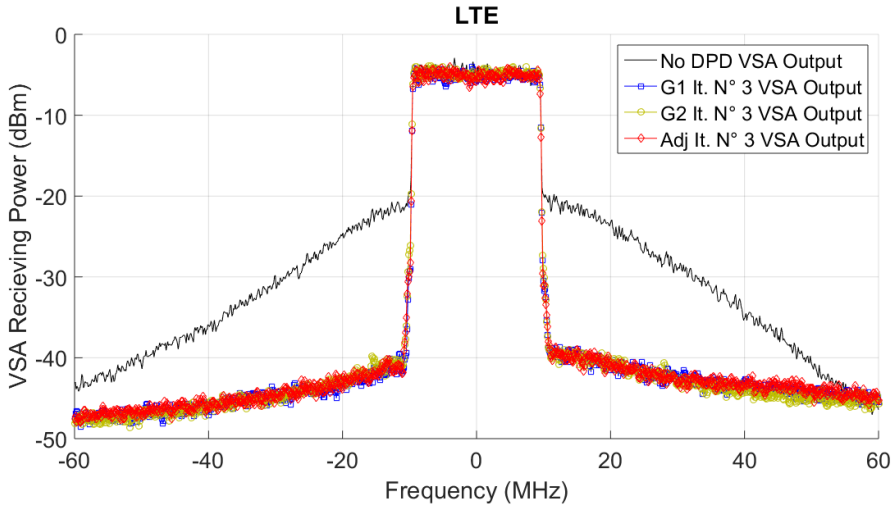



Figure 5.8: Measured spectra of the linearized PA at 45 dBm PA output

### 5.3.3 Performance comparison at 45 dBm & 46 dBm output

In Fig 5.8 and Fig 5.9, the spectra of each linearization method at iteration 3 when PA output average power is 45 dBm and 46 dBm respectively are shown in the figure and compared.

The comparison of the EVM, the ACPR gain and the PAE of 3 linearization methods at the 3rd iteration is shown in Table 5.2. We can see that the 3 linearization methods have very similar performances except the EVM of "G1" is higher. This is due to its instability at the first iteration. The efficiencies are close to that of LTE without DPD in Fig 5.7.

## 5.4 Conclusion

In this chapter, we discussed the impact of PA linear gain choices on linearization performances.

The characterization of PA is an indispensable step for calculating the model of DPD. The characteristics of PA depend on the dynamic power range and the power distribution function of the input signal.

Linearization of PA at different gains are compared with experimental study in terms of ACPR, EVM and PAE. An adjustment approach is very important for PA linearization at $G_1$.

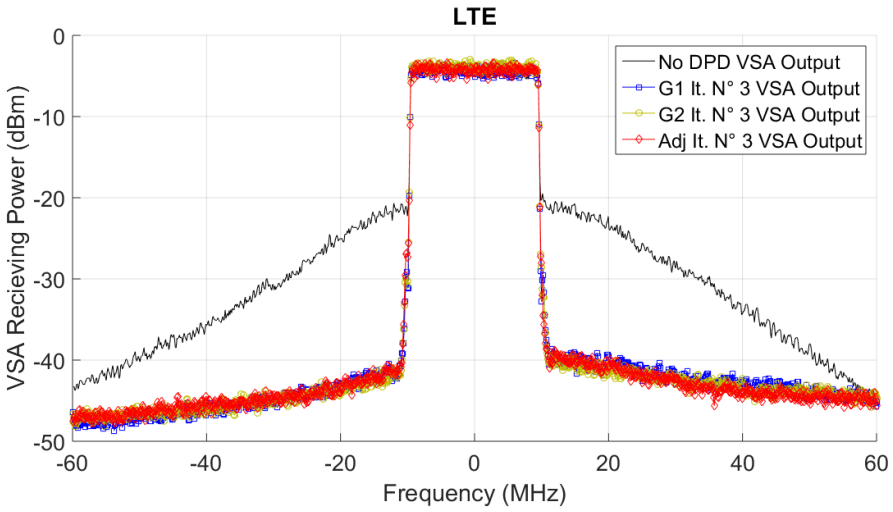



Figure 5.9: Measured spectra of the linearized PA at 46 dBm PA output

Table 5.2: Performance of 3 linearizations

| Output | 45 dBm | | | 46 dBm | | |
|---|---|---|---|---|---|---|
| Method | G1 | Adj | G2 | G1 | Adj | G2 |
| EVM(%) | 7.1 | 3.5 | 5.9 | 8.1 | 5.9 | 7.1 |
| ACPR G L1(dB) | 17.36 | 17.28 | 17.30 | 17.56 | 17.66 | 18.16 |
| PA output(dBm) | 45.24 | 45.24 | 45.35 | 45.81 | 45.95 | 46.3 |
| System Gain(dB) | 44.07 | 44.07 | 41.35 | 43.49 | 43.63 | 41.3 |
| Supply Vol(V) | 28 | 28 | 28 | 28 | 28 | 28 |
| Supply Cur(A) | 3.5 | 3.5 | 3.6 | 4.1 | 4.1 | 4.4 |
| PAE(%) | 34.1 | 34.1 | 34.0 | 33.2 | 34.3 | 34.6 |

Despite the fact that the choice of the nomilization gain is sometimes seen as critical, the experiments conducted in this work for two choices of gain, namely small signal gain and peak power gain, show that:

- their convergences of the linearization performances are similar.
- the efficiency of the linearized PA is not impacted by the chosen gain.

# Chapter 6

# Conclusion and Future Work

In this dissertation, we presented the nonlinearities and memory effects of high power amplifiers and linearized it using digital predistortion. The DPD model structures are studied and implemented on a test bench. This dissertation mainly focuses on study on model structure determination and on cascaded DPD in order to reduce the complexity of DPD models while respecting the linearization performances.

## 6.1 Contributions

The contribution of this dissertation consists of three aspects: Model structure optimization, PA linearization with cascaded DPD, different choices of PA linear gains.

1. For model structure optimization, an algorithm based on hill-climbing heuristic is proposed to search for a model structure which represents a trade-off between the modeling accuracy and model complexity. The proposed algorithm has no randomness comparing with genetic algorithm, which allows more possibilities to optimize the execution time by controlling its search path. With the proposed method, the DPD models can have very few complexity while keeping the linearization performances in a tolerant level.

2. Cascaded DPD are explored to reduce the complexity of model identification, matrix conditioning number and coefficients dynamic range, which have similar performances to a single-block DPD. Matrix condition number has an impact on the accuracy of LS calculation. The coefficients dynamic range can influence the floating-point precision which influences the complexity of implementation

on hardware, e.g. DSP and FPGA.

3. Different choices of PA linear gains are compared. The analysis considers especially their impacts on PAE. An adjustment of DPD model coefficients identification is proposed in the case of small signal gain.

The verification of DPD performances are made on an Ampleon Three-way Doherty PA with output peak power up to 500 W. These measurements have been obtained thanks to the support of National Instruments on Digital Predistortion Framework research activity [53] and the support of Teamcast in the frame of the ambrun project (FUI AAP11) [54].

## 6.2 Future Work

To extend this dissertation, some future work can be done:

1. Adapt the optimal DPD model structure when the input signal is changed. The characteristics of PA depend on many factors, e.g. the signal average power, the probability density function of the signal amplitude, the bandwidth of the signal and the circuit temperature, etc. The optimal DPD needs to be renewed and to fit the characteristics of PA once it is changed. The execution time should be short so that the real-time implementation is viable.

2. Explore the cascaded DPD with more than two stages while each stage have very low complexity. In this dissertation, it is shown that a multi-stage model can have similar linearization performance with a single-stage model. It will be interesting to study the performance of a multi-stage model where each stage has minimum orders (not only a gain) of nonlinearity and memory effect. In another word, the number of basis functions in each stage is limited to four, and we configure the structure by increasing the number of stages. In this case, the complexity of identification is always the minimum.

3. Implement the DPD on digital hardware, e.g. FPGA.

# Bibliography


[1] R. G. P. Colantonio, F. Giannini and L. Piazzon, *Advanced Microwave Circuits and Systems*, V. Zhurbenko, Ed.

[2] J. Vuolevi, T. Rahkonen, and J. Manninen, "Measurement technique for characterizing memory effects in rf power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 49, no. 8, pp. 1383–1389, Aug 2001.

[3] M. Hussein, V. Bohara, and O. Venard, "Multi-stage digital predistortion based on indirect learning architecture," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 6093–6097.

[4] S. Stapleton and F. Costescu, "An adaptive predistorter for a power amplifier based on adjacent channel emissions [mobile communications]," *Vehicular Technology, IEEE Transactions on*, vol. 41, no. 1, pp. 49–56, Feb 1992.

[5] M. G. D. Benedetto and P. Mandarini, "A new analog predistortion criterion with application to high efficiency digital radio links," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2966–2974, Dec 1995.

[6] J. Yi, Y. Yang, M. Park, W. Kang, and B. Kim, "Analog predistortion linearizer for high-power rf amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 12, pp. 2709–2713, Dec 2000.

[7] K.-J. Cho, D.-H. Jang, S.-H. Kim, J.-Y. Kim, J.-H. Kim, and S. P. Stapleton, "An analog compensation method for asymmetric imd characteristics of power amplifier," *IEEE Microwave and Wireless Components Letters*, vol. 14, no. 4, pp. 153–155, April 2004.

[8] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, no. 23, pp. 1417–1418, Nov 2001.

[9] W.-J. Kim, S. Stapleton, J. H. Kim, and C. Edelman, "Digital predistortion linearizes wireless power amplifiers," *Microwave Magazine, IEEE*, vol. 6, no. 3, pp. 54–61, Sept. 2005.

[10] M. Hussein, Y. Wang, B. Feuvrie, S. Toutain, and G. Peyresoubes, "Piecewise complex circular approximation of the inverse characteristics of power amplifiers for digital predistortion techniques," *Digital Telecommunications, 2008. ICDT '08. The Third International Conference on*, pp. 59–63, July 2008.

[11] M. A. Hussein, Y. Wang, G. Peyresoubes, B. Feuvrie, and S. Toutain, "Lut/parametric digital predistortion approach for the linearization of power amplifiers characteristics," *Microwave Conference, 2008. EuMC 2008. 38th European*, pp. 571–574, Oct. 2008.

[12] A. Katz, J. Wood, and D. Chokola, "The evolution of pa linearization: From classic feedforward and feedback through analog and digital predistortion," *IEEE Microwave Magazine*, vol. 17, no. 2, pp. 32–40, Feb 2016.

[13] K. Gumber and M. Rawat, "A modified hybrid rf predistorter linearizer for ultra wideband 5g systems," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[14] F. M. Barradas, L. C. Nunes, T. R. Cunha, P. M. Lavrador, P. M. Cabral, and J. C. Pedro, "Compensation of long-term memory effects on gan hemt-based power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. PP, no. 99, pp. 1–10, 2017.

[15] W. Bosch and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 37, no. 12, pp. 1885–1890, Dec 1989.

[16] S. Boumaiza and F. Ghannouchi, "Thermal memory effects modeling and compensation in rf power amplifiers and predistortion linearizers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 51, no. 12, pp. 2427–2433, Dec. 2003.

[17] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for rf power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 1, pp. 348–359, Jan. 2006.

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

[19] L. Ding, G. Zhou, D. Morgan, Z. Ma, J. Kenney, J. Kim, and C. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *Communications, IEEE Transactions on*, vol. 52, no. 1, pp. 159–165, Jan. 2004.

[20] A. Zhu and T. Brazil, "Behavioral modeling of rf power amplifiers based on pruned volterra series," *Microwave and Wireless Components Letters, IEEE*, vol. 14, no. 12, pp. 563–565, Dec. 2004.

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

[22] P. M. Lavrador, J. C. Pedro, and N. B. Carvalho, "A new volterra series based orthogonal behavioral model for power amplifiers," in *Microwave Conference Proceedings, 2005. APMC 2005. Asia-Pacific Conference Proceedings*, vol. 1, Dec 2005, pp. 4 pp.–.

[23] J. Dooley, B. O'Brien, and T. J. Brazil, "Behavioral modeling of rf power amplifiers using adaptive recursive polynomial functions," in *Microwave Symposium Digest, 2006. IEEE MTT-S International*, June 2006, pp. 852–855.

[24] A. Zhu, J. Pedro, and T. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of rf power amplifiers," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 54, no. 12, pp. 4323–4332, Dec 2006.

[25] S. Benedetto and E. Biglieri, *Principles of Digital Transmission: With Wireless Applications*. Norwell, MA, USA: Kluwer Academic Publishers, 1999.

[26] A. Saleh, "Frequency-independent and frequency-dependent nonlinear models of twt amplifiers," *Communications, IEEE Transactions on*, vol. 29, no. 11, pp. 1715–1720, Nov 1981.

[27] C. Rapp, "Effects of HPA-nonlinearity on a 4-dpsk/ofdm-signal for a digital sound broadcasting system," *Proc. 2nd Eur. Conf. Satellite Communications, Liege, Belgium*, pp. 179–184, Oct. 1991.

[28] S. Zhang, W. Chen, Y.-J. Liu, and F. Ghannouchi, "A time misalignment tolerant 2d-memory polynomials predistorter for concurrent dual-band power amplifiers," *Microwave and Wireless Components Letters, IEEE*, vol. 23, no. 9, pp. 501–503, Sept 2013.

[29] Y.-J. Liu, J. Zhou, W. Chen, and B.-H. Zhou, "A robust augmented complexity-reduced generalized memory polynomial for wideband rf power amplifiers," *Industrial Electronics, IEEE Transactions on*, vol. 61, no. 5, pp. 2389–2401, May 2014.

[30] C. Zhang, J. Wang, and W. Wu, "A piecewise generalized memory polynomial model for envelope tracking power amplifiers," in *2015 Asia-Pacific Microwave Conference (APMC)*, vol. 3, Dec 2015, pp. 1–3.

[31] H. Hemesi, A. Abdipour, and A. Mohammadi, "Analytical modeling of mimo-ofdm system in the presence of nonlinear power amplifier with memory," *IEEE Transactions on Communications*, vol. 61, no. 1, pp. 155–163, January 2013.

[32] A. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 58, no. 6, pp. 1510–1520, June 2010.

[33] R. Raich, H. Qian, and G. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 5, pp. 1468–1479, Sept. 2004.

[34] F. Giri and E.-W. Bai, *Block Oriented Nonlinear System Identification*. Springer, 2010. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01059742

[35] A. Y. Kibangou and G. Favier, "Wiener-hammerstein systems modeling using diagonal volterra kernels coefficients," *IEEE Signal Processing Letters*, vol. 13, no. 6, pp. 381–384, June 2006.

[36] M. Schoukens and K. Tiels, "Identification of nonlinear block-oriented systems starting from linear approximations: A survey," *CoRR*, vol. abs/1607.01217, 2016. [Online]. Available: http://arxiv.org/abs/1607.01217

[37] H. Jiang and P. Wilford, "Digital predistortion for power amplifiers using separable functions," *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4121–4130, Aug 2010.

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

[39] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer Academic Publishers, 1991.

[40] A. Zhu, "Decomposed vector rotation-based behavioral modeling for digital predistortion of rf power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 737–744, Feb 2015.

[41] L. O. Chua and S. M. Kang, "Section-wise piecewise-linear functions: Canonical representation, properties, and applications," *Proceedings of the IEEE*, vol. 65, no. 6, pp. 915–929, June 1977.

[42] J. Zhai, L. Zhang, Z. Yu, J. Zhou, and W. Hong, "A modified canonical piecewise-linear function-based behavioral model for wideband power amplifiers," *IEEE Microwave and Wireless Components Letters*, vol. 26, no. 3, pp. 195–197, March 2016.

[43] Q.-J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial neural networks for rf and microwave design - from theory to practice," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 4, pp. 1339–1350, Apr 2003.

[44] T. Liu, S. Boumaiza, and F. Ghannouchi, "Dynamic behavioral modeling of 3g power amplifiers using real-valued time-delay neural networks," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, no. 3, pp. 1025–1033, March 2004.

[45] M. Rawat and F. Ghannouchi, "Distributed spatiotemporal neural network for nonlinear dynamic transmitter modeling and adaptive digital predistortion," *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 3, pp. 595 –608, march 2012.

[46] M. Abi Hussein, V. Bohara, and O. Venard, "On the system level convergence of ila and dla for digital predistortion," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*, Aug 2012, pp. 870–874.

[47] Y. H. Lim, Y. S. Cho, I. W. Cha, and D. H. Youn, "An adaptive nonlinear prefilter for compensation of distortion in nonlinear systems," *Signal Processing, IEEE Transactions on*, vol. 46, no. 6, pp. 1726–1730, Jun 1998.

[48] C. Eun and E. Powers, "A new volterra predistorter based on the indirect learning architecture," *Signal Processing, IEEE Transactions on*, vol. 45, no. 1, pp. 223–227, Jan 1997.

[49] L. Trefethen and D. Bau III, *Numerical linear algebra*. Society for Industrial Mathematics, 1997, no. 50.

[50] D. Zhou and V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," *Signal Processing, IEEE Transactions on*, vol. 55, no. 1, pp. 120–133, Jan. 2007.

[51] B. Laki and C. Kikkert, "Adaptive digital predistortion for wideband high crest factor applications based on the wacp optimization objective: A conceptual overview," *Broadcasting, IEEE Transactions on*, vol. 58, no. 4, pp. 609 –618, dec. 2012.

[52] R. N. Braithwaite, "Closed-loop digital predistortion (dpd) using an observation path with limited bandwidth," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 726–736, Feb 2015.

[53] M. Hussein, S.Wang, G.Baudoin, and O.Venard, "Framework for rf pa linearization by adaptive digital predistortion," EuMW 2015, NI-AWR RF/Microwave PA Forum, Sept 2015.

[54] Teamcast. (2015) Ambrun project (fui aap11). [Online]. Available: http://www.teamcast.com/innovation/reseach-and-development/

[55] H. Jiang, X. Yu, and P. Wilford, "Digital predistortion using stochastic conjugate gradient method," *Broadcasting, IEEE Transactions on*, vol. 58, no. 1, pp. 114 –124, march 2012.

[56] X. Yu and H. Jiang, "Digital predistortion using adaptive basis functions," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 60, no. 12, pp. 3317–3327, Dec 2013.

[57] R. N. Braithwaite, "Pruning strategies for a volterra series model used in digital predistortion (dpd) of rf power amplifiers," in *2017 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, Jan 2017, pp. 4–7.

[58] H. Jiang, G. Huang, P. A. Wilford, and L. Yu, "Constrained and preconditioned stochastic gradient method," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2678–2691, May 2015.

[59] W. Chen, S. Zhang, Y. J. Liu, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Efficient pruning technique of memory polynomial models suitable for pa behavioral modeling and digital predistortion," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 10, pp. 2290–2299, Oct 2014.

[60] J. Reina-Tosina, M. Allegue-Martinez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 63, no. 2, pp. 745–753, Feb 2015.

[61] E. Zenteno, Z. A. Khan, M. Isaksson, and P. Händel, "Finding structural information about rf power amplifiers using an orthogonal nonparametric kernel smoothing estimator," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2883–2889, May 2016.

[62] T. Gotthans, G. Baudoin, and A. Mbaye, "Optimal order estimation for modeling and predistortion of power amplifiers," in *Microwaves, Communications, Antennas and Electronics Systems (COMCAS), 2013 IEEE International Conference on*, Oct 2013, pp. 1–4.

[63] A. Abdelhafiz, O. Hammi, A. Zerguine, A. Al-Awami, and F. Ghannouchi, "A pso based memory polynomial predistorter with embedded dimension

estimation," *Broadcasting, IEEE Transactions on*, vol. 59, no. 4, pp. 665–673, Dec 2013.

[64] K. Freiberger, M. Wolkerstorfer, H. Enzinger, and C. Vogel, "Digital predistorter identification based on constrained multi-objective optimization of wlan standard performance metrics," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015, pp. 862–865.

[65] S. Wang, M. A. Hussein, O. Venard, and G. Baudoin, "Optimal sizing of generalized memory polynomial model structure based on hill-climbing heuristic," in *Microwave Conference (EuMC), 2016 European*, Oct 2016, accepted.

[66] M. Younes, O. Hammi, A. Kwan, and F. Ghannouchi, "An accurate complexity-reduced "plume" model for behavioral modeling and digital predistortion of rf power amplifiers," *Industrial Electronics, IEEE Transactions on*, vol. 58, no. 4, pp. 1397–1405, April 2011.

[67] M. Rawat, K. Rawat, F. Ghannouchi, S. Bhattacharjee, and H. Leung, "Generalized rational functions for reduced-complexity behavioral modeling and digital predistortion of broadband wireless transmitters," *Instrumentation and Measurement, IEEE Transactions on*, vol. 63, no. 2, pp. 485–498, Feb 2014.

[68] F. Mkadem, M. C. Fares, S. Boumaiza, and J. Wood, "Complexity-reduced volterra series model for power amplifier digital predistortion," *Analog Integrated Circuits and Signal Processing*, vol. 79, no. 2, pp. 331–343, 2014.

[69] K. Deh, S. Prasad, P. Colantonio, and G. Rocco, "Microwave circuit models using the structured genetic algorithm," in *TELSIKS 2005 - 2005 uth International Conference on Telecommunication in ModernSatellite, Cable and Broadcasting Services*, vol. 2, Sept 2005, pp. 367–372 vol. 2.

[70] J. Sills and R. Sperlich, "Adaptive power amplifier linearization by digital pre-distortion using genetic algorithms," *Radio and Wireless Conference, 2002. RAWCON 2002. IEEE*, pp. 229–232, 2002.

[71] C. Çiflikli and A. Ç. Yapıcı, "Genetic algorithm optimization of a hybrid analog/digital predistorter for rf power amplifiers," *Analog Integrated Circuits and Signal Processing*, vol. 52, no. 1, pp. 25–30, 2007. [Online]. Available: http://dx.doi.org/10.1007/s10470-007-9100-6

[72] B. Liu, Y. Wang, Z. Yu, L. Liu, M. Li, Z. Wang, J. Lu, and F. V. Fernández, "Analog circuit optimization system based on hybrid evolutionary algorithms," *Integration, the {VLSI} Journal*, vol. 42, no. 2, pp. 137 – 148, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167926008000126

[73] E. Tlelo-Cuautle, I. Guerra-Gómez, L. G. de la Fraga, G. Flores-Becerra, S. Polanco-Martagón, M. Fakhfakh, C. A. Reyes-García, G. Rodríguez-Gómez, and G. Reyes-Salgado, *Evolutionary Algorithms in the Optimal Sizing of Analog Circuits*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 109–138. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21705-0_5

[74] S. Mallick, R. Kar, D. Mandal, and S. P. Ghoshal, "Optimal sizing of cmos analog circuits using gravitational search algorithm with particle swarm optimization," *International Journal of Machine Learning and Cybernetics*, pp. 1–23, 2015. [Online]. Available: http://dx.doi.org/10.1007/s13042-014-0324-3

[75] L. Li, A. Ghazi, J. Boutellier, L. Anttila, M. Valkama, and S. S. Bhattacharyya, *Evolutionary Multiobjective Optimization for Digital Predistortion Architectures*. Cham: Springer International Publishing, 2016, pp. 498–510. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-40352-6_41

[76] K. Sullivan and S. Jacobson, "A convergence analysis of generalized hill climbing algorithms," *Automatic Control, IEEE Transactions on*, vol. 46, no. 8, pp. 1288–1293, Aug 2001.

[77] L. Rueda and V. Vidyadharan, "A hill-climbing approach for automatic gridding of cdna microarray images," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 3, no. 1, pp. 72–83, Jan 2006.

[78] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Prentice Hall, 2009.

[79] W. Greblicki and M. Pawlak, "Nonparametric identification of a particular nonlinear time series system," *Signal Processing, IEEE Transactions on*, vol. 40, no. 4, pp. 985–989, Apr 1992.

[80] R. Braithwaite and S. Carichner, "An improved doherty amplifier using cascaded digital predistortion and digital gate voltage enhancement," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 57, no. 12, pp. 3118–3126, Dec 2009.

[81] M. Rawat, F. Ghannouchi, and K. Rawat, "Three-layered biased memory polynomial for dynamic modeling and predistortion of transmitters with memory," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, no. 3, pp. 768–777, March 2013.

[82] S. Chen, "An efficient predistorter design for compensating nonlinear memory high power amplifiers," *Broadcasting, IEEE Transactions on*, vol. 57, no. 4, pp. 856 –865, dec. 2011.

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

[84] O. Hammi, M. S. Sharawi, and F. M. Ghannouchi, "Generalized twin-nonlinear two-box digital predistorter for gan based lte doherty power amplifiers with strong memory effects," in *2013 IEEE International Wireless Symposium (IWS)*, April 2013, pp. 1–4.

[85] O. Hammi, A. Kwan, S. Bensmida, K. A. Morris, and F. M. Ghannouchi, "A digital predistortion system with extended correction bandwidth with application to lte-a nonlinear power amplifiers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 12, pp. 3487–3495, Dec 2014.

[86] B. F. Beidas, "Adaptive digital signal predistortion for nonlinear communication systems using successive methods," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2166–2175, May 2016.

[87] J. Qiu, D. Abe, T. Antonsen, B. Danly, B. Levush, and R. Myers, "Higher-order predistortion linearisation using cascaded lower-order nonlinearities," *Electronics Letters*, vol. 40, no. 17, pp. 1068–1070, Aug 2004.

[88] O. Hammi and F. M. Ghannouchi, "Twin nonlinear two-box models for power amplifiers and transmitters exhibiting memory effects with applica-

tion to digital predistortion," *IEEE Microwave and Wireless Components Letters*, vol. 19, no. 8, pp. 530–532, Aug 2009.

[89] K. Narendra and P. Gallman, "An iterative method for the identification of nonlinear systems using a hammerstein model," *Automatic Control, IEEE Transactions on*, vol. 11, no. 3, pp. 546–550, Jul 1966.

[90] F. Chang and R. Luus, "A noniterative method for identification using hammerstein model," *Automatic Control, IEEE Transactions on*, vol. 16, no. 5, pp. 464–468, Oct 1971.

[91] S. Rangan, G. Wolodkin, and K. Poolla, "New results for hammerstein system identification," in *Proceedings of 1995 34th IEEE Conference on Decision and Control*, vol. 1, Dec 1995, pp. 697–702 vol.1.

[92] A. Wills, T. B. Schön, L. Ljung, and B. Ninness, "Identification of hammerstein–wiener models," *Automatica*, vol. 49, no. 1, pp. 70 – 81, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109812004815

[93] L. Ding, R. Raich, and G. Zhou, "A hammerstein predistortion linearization design based on the indirect learning architecture," *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, vol. 3, pp. III–2689–III–2692 vol.3, 2002.

[94] W. Greblicki and M. Pawlak, "Identification of discrete hammerstein systems using kernel regression estimates," *Automatic Control, IEEE Transactions on*, vol. 31, no. 1, pp. 74–77, Jan 1986.

[95] E. A. Nadaraya, "On estimating regression," *Theory of Probability & Its Applications*, vol. 9, no. 1, pp. 141–142, 1964. [Online]. Available: https://doi.org/10.1137/1109020

[96] G. S. Watson, "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 26, no. 4, pp. 359–372, 1964. [Online]. Available: http://www.jstor.org/stable/25049340

[97] W. Greblicki, "Nonparametric identification of wiener systems," *Information Theory, IEEE Transactions on*, vol. 38, no. 5, pp. 1487–1493, Sep 1992.

[98] E.-W. Bai and D. Li, "Convergence of the iterative hammerstein system identification algorithm," *Automatic Control, IEEE Transactions on*, vol. 49, no. 11, pp. 1929–1940, Nov 2004.

[99] L. Ljung, "Estimating linear time-invariant models of nonlinear time-varying systems," *European Journal of Control*, vol. 7, no. 2, pp. 203 – 219, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S094735800171148X

[100] J. Schoukens, R. Pintelon, Y. Rolain, M. Schoukens, K. Tiels, L. Vanbeylen, A. V. Mulders, and G. Vandersteen, "Structure discrimination in block-oriented models using linear approximations: A theoretic framework," *Automatica*, vol. 53, no. Supplement C, pp. 225 – 234, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109814006244

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

[102] K. Muhonen, M. Kavehrad, and R. Krishnamoorthy, "Look-up table techniques for adaptive digital predistortion: a development and comparison," *Vehicular Technology, IEEE Transactions on*, vol. 49, no. 5, pp. 1995–2002, Sep 2000.

[103] A. Zhu, P. J. Draxler, J. J. Yan, T. J. Brazil, D. F. Kimball, and P. M. Asbeck, "Open-loop digital predistorter for rf power amplifiers using dynamic deviation reduction-based volterra series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 7, pp. 1524–1534, July 2008.

[104] H. Wang, F. Liu, W. Tao, and G. Li, "One-step extraction of optimal normalisation gain for digital predistortion linearisation," *Electronics Letters*, vol. 51, no. 6, pp. 514–516, 2015.

[105] L. Aladren, P. Garcia-Ducar, P. L. Carro, and J. de Mingo, "Digital predistortion optimization using normalization gain adjustment in wideband systems," in *2013 European Microwave Conference*, Oct 2013, pp. 420–423.

[106] H. Qian, H. Huang, and S. Yao, "A general adaptive digital predistortion architecture for stand-alone rf power amplifiers," *IEEE Transactions on Broadcasting*, vol. 59, no. 3, pp. 528–538, Sept 2013.

[107] J. Chani-Cahuana, C. Fager, and T. Eriksson, "A new variant of the indirect learning architecture for the linearization of power amplifiers," in *2015 European Microwave Conference (EuMC)*, Sept 2015, pp. 1295–1298.