TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

DISSERTATION

Black-Box Modeling of Microwave Amplifiers for Linearization

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

M. Sc. Daniel D. Silveira
Matr.Nr. 0326822
Guglgasse 8/2/3/9D
A-1110 Wien
geb. 20. Jänner 1978

Wien, 28. November 2007

Begutachter:

O. Univ. Prof. Dipl.-Ing. Dr. techn. Gottfried Magerl
Institut für Elektrische Mess- und Schaltungstechnik
Technische Universität Wien — Österreich

Prof. Dr. Eduard Bertran Albertí
Control, Monitoring and Communications Group
Technical University of Catalonia — Spain

Abstrakt

In dieser Dissertation werden Blackbox Modelle von Mikrowellen-Leistungsverstärkern vorgestellt und untersucht.

Blackbox Modelle oder Verhaltensmodelle sind Teil von theoretischen und experimentellen Methoden der Systemtheorie, und sind in der Systemidentifikation angewendet. Die Modelle werden aus eingangs- und ausgangsseitigen Messdaten generiert, ohne die innere Struktur der zu modellierenden Einheit zu kennen. Hauptaugenmerk liegt dabei auf der Struktur dieser Modelle, den Schätzverfahren für die Modellparameter, Evaluierungsmethoden und Entwicklung neuer Modelle.

Praktische Einsatzmöglichkeiten der Verhaltensmodelle von Leistungsverstärkern sind beispielsweise in Programmen zur Schaltungssimulation und in Linearisierungsverfahren zu finden. Die somit erhaltenen Modellparameter sind dabei repräsentativ für das Verhalten der physikalischen Einheit. Leistungsverstärker sind jedoch nichtlineare, gedächtnisbehaftete Bauelemente (d. h., das aktuelle Ausgangssignal von dem gegenwärtigen und vergangenen Eingagssignal abhängt). Daher ist es eine besonders anspruchsvolle Aufgabe, ein Modell zu extrahieren, welches das (nichtlineare) Verhalten des Leistungsverstärkers in ausreichend hoher Qualität beschreibt, und dabei auch Speichereffekte berücksichtig.

Der erste Teil dieser Dissertation beschäftigt sich mit einer Analyse bezüglich jener Signale, die zur Anregung von Verhaltensmodellen geeignet sind, sowie Unterteilung von Daten in spezielle Datensätze und der Auswertung von Bewertungskriterien zur Feststellung der Modellgüte. Außerdem werden lineare Schätzverfahren und Parametrisierung linearer Systeme untersucht, wobei die Fortschritte von Schätzverfahren mit FIR Filtern aufgezeigt werden. Zusätzlich zu den linearen Schätzverfahren werden auch Methoden zur nichtlinearen Systemidentifikation, insbesondere von Mikrowellen-Leistungsverstärkern, vorgestellt. Hervorgehoben werden statische und dynamische nichtlineare Modell, zusammen mit den zugehörigen, linearen Schätzmethoden.

Der zweite Beitrag dieser Dissertation ist eine Studie von ausgewählten Anwendungen von Verhaltensmodellen. Abschnitt 5.3 evaluiert Verhaltensmodelle von Leistungsverstärkern, derer Modellparameter mit verschiedenen Eingangsrauschpegeln geschätzt worden sind. Dabei stellt sich heraus, daßdie Modelle, die mit vorhandenem Eingangsrauschen extrahiert worden sind, in der Regel bessere Resultate liefern als jene, die ohne Eingangsrauschen bestimmt worden sind. Jedoch verhalten sich Letztere optimal im rauschfreien Betrieb.

Die nächste Anwendung, die in Abschnitt 5.4 vorgestellt wird, ist die Implementierung einer Vorverzerrung in einem Mikrowellen-Leistungsverstrker. Der Vorverzerrer lieferte annehmbare Resultate, welche durch detaillierter ausgearbeitete Modelle verbessert werden konnten.

Im Abschnitt 5.5 werden Fortschritte in der Verhaltensmodellierung aufgezeigt. Beispielweise wird eine Volterra Modellapproximation (Wiener-Bose) vorgestellt, wel-

ches gut strukturiert ist und lineare Schätzmethode verwendet. Das Hauptproblem der Schätzverfahren zur Volterra Approximation ist die schlechte konditionierte Hessianmatrix. Hier wird eine Methode zur Verbesserung der Konditionszahl von der Hessianmatrix dargestellt, basierende auf der Änderung der Abtastrate, die in der Regressionmatrix angewandt wurde.

In weiterer Folge wird ein Bewertungskriterium entwickelt, das zur Evaluierung der Modellierung von Nichtlinearitäten am Ausgang des Leistungsverstärkers herangezogen wird. Das Kriterium wurde an gedächtnislosen, linear gedächtnisbehafteten, nichtlinear gedächtnisbehafteten, sowie an linear und nichtlinear gedächtnisbehafteten Systemen angewandt. Die Ergebnisse werden in Abschnitt 5.6 präsentiert.

Eine Vorverarbeitung für die Parametrisierung von Verhaltensmodellen für Leistungsverstärker wird in Abschnitt 5.7 präsentiert. Eine Klasse von Modellen basierend Zweige zusammengesetzt aus einer Lookup-Tabelle gefolgt von einen ausgedünnten Volterra Model (reduzierte Wiener-Bose Struktur) wird vorgestellt. Unter Verwendung von Multiraten Signalverarbeitung im Identifikationsprozess erhält man hochpräzise Modelle mit paralleler Struktur. Dieser Zugang erlaubt eine effizientere Auswertung im Vergleich zu Modellen mit einem einzelnen Zweig und einer dementsprechend höheren Anzahl an Parametern. Die erzielten Ergebnisse bestätigen die Leistungsfähigkeit dieser Modelle das Verhalten eines Leistungsverstärkers zu verkörpern und dabei höhere Genauigkeiten zu erzielen als die zuvor analysierten Modelle. Simulations- und/oder Messdaten bildeten die Grundlage für alle präsentierten Modellierungs- und Validierungsergebnisse.

Im letzen Kapitel wird ein Fazit der durchgeführten Arbeiten gezogen und auf öffene Aspekte hingewiesen welchen der Kernpunkt weitererführender Forschungsaktivitäten sein werden.

Abstract

This thesis is a research about PA BMs structures and their estimation strategies for digital pre-distortion purposes.

Black-box or behavioral models (BM) are obtained from input/output observations of a system, without knowledge of its inner structure. They can be optimized for a specific system, and so it is possible to represent the physical component behavior by this model (e.g., RF power amplifiers – PAs).

The principal applications of PA BMs are linearization and circuit simulation tools. PAs are nonlinear devices with memory effects (i. e., the actual output signal value depends of the present and past input signal values), therefore, it is a challenging task to extract their equivalent BM.

The work starts with a brief overview of BMs excitation signals and partitioning of data used in the modeling process. Figures of merit, tools to measure BMs quality, are analyzed.

An investigation of linear estimation techniques and parametrization of linear systems is also performed, showing advances in the finite impulse response filter estimation.

Following, techniques for nonlinear systems estimation are described, focused on PAs. Static nonlinear models and dynamic ones are outlined, together with their linear estimation methods.

In the second part of this work, selected applications of behavioral models are surveyed.

In Section 5.3, a study about PA BMs estimated under different noise levels shows that models estimated with noise corrupted data achieve better general results than models estimated with noise free data. Models estimated using noise free data have an optimal performance only if noise free data is used.

The next application was a PA pre-distorter implementation, in Section 5.4. The pre-distorter presented reasonable results, that could be enhanced by the use of more elaborated models.

In Section 5.5, advances in BMs are shown, as a Volterra series model approximation (Wiener-Bose) that uses linear estimation techniques, having a well organized structure. The main problem when estimating Volterra approximation models is the ill-conditioned Hessian. A method for improving the condition number of the least-squares Hessian is depicted, based on a resampling factor used in the regression matrix.

A figure of merit was developed to analyze the modeling of nonlinear distortions of the output signal is shown in Section 5.6. Different models presented better performance for systems with specific memory effects (memoryless, linear memory, nonlinear memory, linear and nonlinear memory). Models have to be optimized for each situation to avoid noise modeling.

A pre-processor for PA BMs estimation is presented in Section 5.7. New models are introduced, composed by branches with look-up tables and pruned Volterra series

approximation (Wiener-Bose reduced). Using a resampling factor in the fitting process, these models have high accuracy and operate in parallel, what is more computationally efficient than models having a single branch with a high number of parameters. The results have shown that these parallel models were more accurate and capable to a better representation of a PA than previously analyzed and developed models.

Simulated and/or measured data are used for modeling and validation in all cases.

Conclusions about the work and future research are drawn in the last chapter.

To João, Regina and Helena.
Thanks for your
unconditional help.

Acknowledgments

Many people contributed to this work, directly or indirectly. It's a hard task to list everybody.

Special thanks to:

- My supervisor, Gottfried Magerl, always prompt to help, as well as my co-supervisor, Michael Gadringer;
- Holger Arthaber and Ayse Adalan, for their valuable revision of this work;
- Markus Mayer, Christian Schubert, Pere Gilabert, Gabriel Montoro, and all EMST personal for their help in several issues;
- Prof. Eduard Bertran and Prof. José Carlos Pedro for fruitful discussions and ideas;
- My family and friends;
- CAPES–Centro de Aperfeicoamento do Ensino Superior, Brazil, for the scholarship that made everything possible; Vienna University of Technology, for accepting me as a student, providing tools and resources to complete my studies; and TARGET–Top Amplifier Research Group in an European Team, a Network of Excellence, supported by the Information Society Technologies Programme of the EU under contract IST-1-507893-NOE, www.target-net.org, which gave me the opportunity to know other researches in different universities and countries, get trained in new technologies, and interchange ideas with highly qualified professionals.

Contents

List of Figures

List of Tables

Abbreviations

%VAF *Variance Accounted For in percentage*

ACPR *Adjacent Channel Power Ratio*

ACLR *Adjacent Channel Leakage Ratio*

ASYM *Asymmetric Responses to Symmetric Input Signal Changes*

AWGN *Additive White Gaussian Noise*

BLA *Best Linear Approximation*

BM *Behavioral Model*

CCDF *Complementary Cumulative Distribution Function*

CF *Crest Factor*

CHAOS *Highly Irregular Responses to Simple Inputs Like Impulses, Steps or Sinusoids*

CN *Condition Number*

DPD *Digital Pre-Distortion*

EVM *Error Vector Magnitude*

FIR *Finite Impulse Response*

FOM *Figure of Merit*

HARM *Generation of Higher-Order Harmonics in Response to a Sinusoidal Input*

IBO	*Input Back-Off*
IDS	*Input Dependent Stability*
IIR	*Infinite Impulse Response*
IRF	*Impulse Response Function*
IM	*Input Multiplicity*
IMD	*Intermodulation Distortion*
IRF	*Impulse Response Function*
LM	*Linear Memory*
LMS	*Least Mean Squares*
LS	*Linear Least Squares*
LUT	*Look-Up Table*
MDL	*Minimum Descriptor Length*
NCD	*Noise Corrupted Data*
NFD	*Noise Free Data*
NLM	*Nonlinear Memory*
NMSE	*Normalized Mean Square Error*
NN	*Neural Network*
noM	*Memoryless PA Power Amplifier*
OM	*Output Multiplicity*
PA	*Power Amplifier*
PAPR	*Peak-to-Average Power Ratio*
PSA	*Performance Spectrum Analyzer*

PSD	*Power Spectral Density*
QAM	*Quadrature Amplitude Modulation*
RF	*Radio Frequency*
RLS	*Recursive Least Squares*
SHAM	*Generation of Sub-Harmonics in Response to any Periodic Input*
SNR	*Signal-to-Noise Ratio*
SVD	*Singular Value Decomposition*
TLS	*Total Least Squares*
VS	*Volterra Series*
WBose	*Wiener-Bose Model*
WCDMA	*Wide-band Code Division Multiple Access*

List of Symbols

$\boldsymbol{\theta}$	*Nonlinear system parameter vector*
$\mathbf{A}$	*Matrix*
$\mathbf{a}$	*Vector*
$\mathbf{a}_k$	*Vector, index k*
$a(k)$	*Vector component*
c_p	p^{th} *order nonlinear coefficient*
$\mathrm{diag}(\cdot)$	*Main diagonal of a matrix*
$\mathbf{h}$	*Linear system parameter vector*
$h(\tau)$	*Impulse response function – linear systems*
$h_p(\cdot)$	p^{th} *order Volterra kernel*
M	*Number of filter delay taps*
N	*Number of observation points*
$\mathbf{U}$	*Least squares regression matrix*
$\mathbf{u}_k$	*Input observation signal (vector)*
$var(\cdot)$	*Variance*
$\mathbf{W}$	*Weight matrix for least squares estimation*
$\mathbf{y}_k$	*Output observation signal (vector)*
$Y(f)$	*Fourier transform of a signal*

$\|\|\cdot\|\|_F$	*Frobenius norm*
$< \cdot >$	*Expectation operator*
$\lfloor \cdot \rfloor$	*Floor operator*
$[\mathbf{U}\|\mathbf{y}]$	$\boldsymbol{U}$ *and* $\boldsymbol{y}$ *are stacked side-by-side*

Chapter 1

Introduction

Modern wireless communications systems use multilevel and/or multi-carrier modulation formats in order to obtain high transmission rates in the assigned bandwidth, requiring bandwidth efficiency, since frequency resources are always limited. Those efficient modulation formats (presenting high peak to average power ratios, PAPR) are also very sensitive to the inter-modulation distortion (IMD), that results from mild nonlinearities in the RF transmitter chain (see Fig. 1.1). All these issues demand power amplifier (PA) operation at significant back-off levels, thus power is used inefficiently.

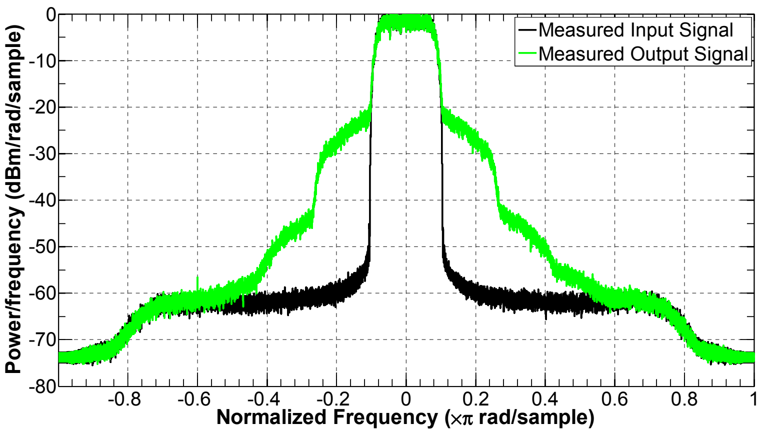

Figure 1.1: *Example of IMD caused by a PA in a 16-QAM modulated signal.*

In addition, modern communication standards are thought to operate with high data rates, where memory effects (i. e., the actual output signal value depends on the present and past input signal values) in PAs cannot be ignored. The PA is not only one of the most power consuming components in the transmitter, but also the cause of the main nonlinear effects in the transmitter chain.

In summary, a requirement of actual communication systems is a highly linear and efficient PA. The main reasons are:

- **Cost of the amplifier:** Using cheaper transistors operating at a higher efficiency will reduce the overall system price;
- **Digital modulation:** With the increase of bits/symbol and higher data rates, signals to be amplified have a high crest factor (CF), forcing the amplifier to

operate in significant back-off in order to reduce the distortion and achieve the desired error vector magnitude (EVM);

- **Power consumption:** A lower consumption is crucial for all systems, from battery powered devices to high power base station amplifiers.

A well designed amplifier can meet some of the above cited requirements. However, but only when the signal is pre-distorted satisfactory levels of efficiency can be achieved. In this sense, the need of accurate pre-distorters can improve the IMD cancelation, as shown in Fig. 1.2. This figure shows that not only the magnitude but also the phase of the models used in the pre-distortion process have to be well determined to achieve good cancelation results.

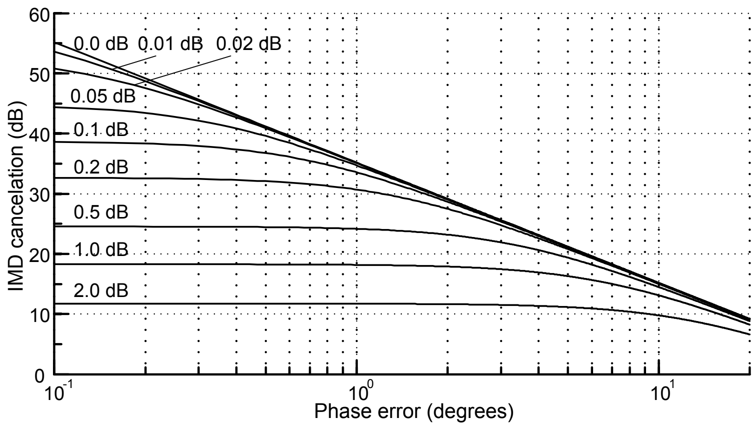

Figure 1.2: *IMD cancelation that can be obtained from a basic pre-distorter for different values of gain and phase error [1].*

The most accurate models are based on input/output measurements of the PA (behavioral models). They are used for pre-distortion purposes but are of limited use in the design of a new PA [2]. For this task, circuit equivalent models, based on semiconductors physics, are employed. These models are characterized by the high number of parameters and the difficult fitting. The accuracy is lower than with behavioral models, so their use in pre-distortion circuits is restricted.

Motivation and Problem Formulation

An actual problem in power amplifier (PA) behavioral models (BMs) research area is that a large number of them are reported in the literature, but unfortunately many authors do not present the extraction procedure in sufficient detail to permit its implementation, or use computationally intensive nonlinear estimation techniques. Many works do not perform a comparison with other models in a way that it is possible to distinguish differences, strengths and weakness of each model. A reasonable use of BMs is only possible if they have a well-organized structure and a moderate complexity of the estimators used, so their characteristics could be completely understood.

This thesis is about BMs for PAs, designed for complex-valued signals. The main objective is the investigation of a subset of existing baseband time-domain BMs, comparing their performance and possible estimation methods, and to propose new models and estimation techniques, using as much previous information as possible. Its motivation is the implementation of efficient BMs, using the simplest estimation techniques available, thus being computationally efficient. These new models find applications mainly in linearization [3] and system simulation tools, as e. g. Advanced Design Systems from Agilent, as commented below.

Behavioral models

Behavioral models are employed in different areas of science, as automotive, medicine, geology, and control engineering. They do not need an a priori knowledge of the system internal composition, for that reason they are also known as black-box models [4]. The objective is to derive a mathematical model from input/output observations that describes the passive or active system behavior with good accuracy (see Fig. 1.3). Therefore, their exactness is highly sensitive to the adopted model structure and the parameter extraction procedure (optimization algorithm).

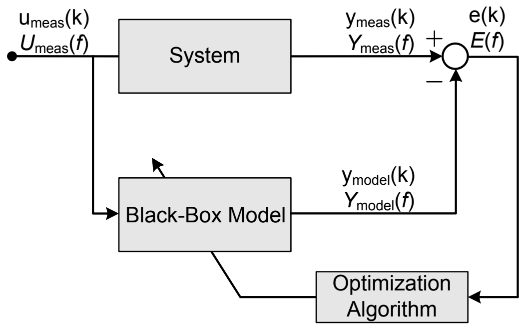

Figure 1.3: *A general modeling procedure.*

The input/output measurement signals and the error signal in the (sampled) time domain are denoted by $u_{meas}(k)$, $y_{meas}(k)$, and $e(k)$, respectively. In the frequency domain they are represented by $U_{meas}(f)$, $Y_{meas}(f)$, and $E(f)$.

Adding some a-priori information to the estimation process can always improve results, instead of doing a "blind estimation". So, a good knowledge of the system to be modeled is necessary in order to advantageously restrict the possibilities of model implementation.

Linearization:

In digital pre-distortion (DPD) linearizers an accurate reproduction of PA memory is of particular importance for a later compensation of PA nonlinear dynamics [5]. The extraction of PA BMs for DPD linearization purposes is carried out by means of input/output complex envelope signal observations.

Figure 1.4 shows two possible architectures for DPD purposes (direct and indirect learning) [6]. In the direct learning case, the model is calculated and inverted to minimize the distortion. In indirect learning, the inverse amplifier BM is calculated using the attenuated output signal and the input signal after pre-distortion. This BM is then used to pre-distort the signal. In both cases, the BM structure and extraction strategy is crucial for accurate results.

Application examples of the direct and indirect learning approaches are given in [7–10].

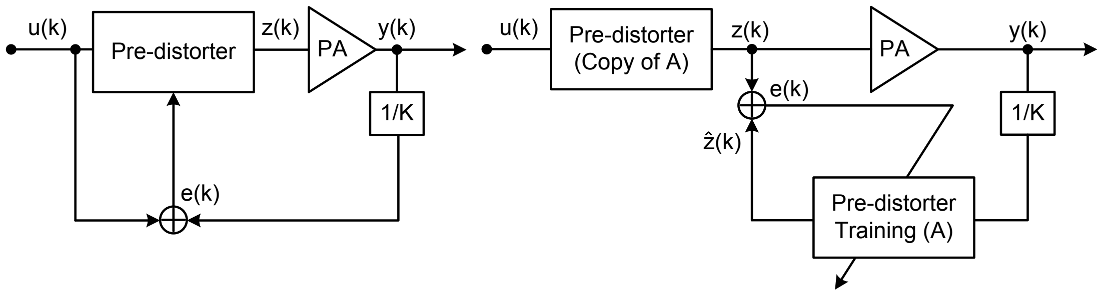

Figure 1.4: *Direct (left) and indirect (right) learning architecture for pre-distorters.*

System Simulation:

In system simulations, an entire amplifier can be represented by a BM. This procedure protects the company's intellectual property, as internal transistor effects are represented with reasonable accuracy. The model is totally described by its coefficients [11]. However, these extracted parameters have not a direct relationship to physical parameters within a PA. Indeed, these models deliver results very close to reality.

1.1 Thesis Outline and Contributions

The thesis is organized as:

Chapter 1: Introduction on BMs and applications, thesis problem formulation and organization, listing of contributions, and a review of main bibliographies in the area.

Chapter 2: BMs particularities, as excitation signals, and organization of observation data. An investigation about figures of merit (FOMs) showing their advantages/disadvantages is presented.

Chapter 3: Linear estimation techniques, and how they are applied for linear systems as finite impulse response (FIR) filter estimation. A contribution here was the adaptation and the use of singular value decomposition (SVD) techniques [12] for FIR filters estimation operating with complex-valued signals [13]. The advantages are highlighted. A FIR frequency-domain estimation technique is also shown.

Chapter 4: A classification of nonlinear systems is depicted, with the main focus on PA. Following, several PA models are analyzed. The static models covered are: power series, baseband power series using orthogonal polynomials, Saleh model, and neural networks configured only with tangent-sigmoid basis functions. Also a pre-processing technique for estimation of baseband power series is presented. The dynamic models covered are: Wiener, Hammerstein, and their relation to Volterra series; parallel configured models, and Wiener-Bose model – a direct approximation of Volterra series. Contributions here were the application of SVD techniques to nonlinear models, and an analysis about Volterra kernels of PAs [14]. A comparison among several models (including some of these models commented) from different universities was coordinated and presented in [15].

Chapter 5: This chapter is about recent developments in BMs. It includes identification in the presence of different levels of additive white Gaussian noise (AWGN) [13], and a predistorter based on Wiener and Hammerstein models [16]. Further on an introduction of the Wiener-Bose model is presented: a Volterra series approximation using linear techniques [17]. It is followed by an alternative FOM for nonlinear distortion identification. Finally, a discussion about parallel models, a pre-processing for PA BM estimation, and a parallel model composed of the combination of the Wiener-Bose model and look-up tables (LUTs) are presented in the last section.

Chapter 6: Draws the final conclusions and discusses some possible and recommendable future research directions.

1.2 Bibliography Review

This section will present some of the main references in the area, but is away from being exhaustive.

The basis theory for nonlinear systems, including a wide description is given in [18–20]. A modern and extensive theoretical overview of BMs for PAs is presented in [21]. In [22] another overview is reported, oriented to automation and control area, and in [23] a comprehensive description of linear and nonlinear systems is encountered, oriented to physiological systems, as well as in [24], which has a practitioner's guide. A complete classification of nonlinear systems is reported in [25]. The theory about baseband polynomials is well explained in [26, 27]. The main studies on orthogonal polynomials for baseband complex-valued signals are in [28, 29]. A discussion about Wiener models, Hammerstein models, and parallel structures, as well as figures of merit, is in [30–33]. A recursive BM is found in [34]. An interesting reference when dealing with Volterra approximations combined with adaptive filter techniques is [35]. References [36,37] present advanced studies about Volterra series (VS) used in modeling of PAs and pruning techniques for VS. Linear and nonlinear methods for estimating BMs are in [38–40]. Information about neural network structures and estimation is given in [41, 42]. Pre-distortion techniques using BMs can be found in [3, 10, 43].

Chapter 2

System Identification – Overview

2.1 Introduction

This chapter will describe excitation signals, measurement data partitioning, and frequently used figures of merit (FOMs).

2.2 Excitation Signal

The selection of an excitation signal is essential to build an efficient model. The ideal input signal for system identification is white Gaussian noise [18, 24]. It excites every nonlinearity source of an unknown system, but it is hard to generate, due to the wide bandwidth and randomness required.

In the ideal case, this signal would have an impulse as auto-correlation function. But, for best results, it is necessary that the signal maintain these characteristics also for higher order correlations with its delayed version. The reasons are explained in sections 3.2 and 5.5.4.

In practice, signals are generated as close to Gaussian noise as possible, using modulated signals. In the frequency domain, they are limited to the system bandwidth and measured up to the highest IMD product to be analyzed.

2.3 Measurement Data Partitioning

The measurement data used in the modeling process have to be partitioned into [40]:

- **Training Data**: Used for the estimation process;
- **Validation Data**: Used to validate the model. Should be different from the training data;
- **Test Data**: Obtained from another realization or a third part of the measurement data. Used to verify the model quality.

A graphical representation of these different types of data is shown in Fig. 2.1.

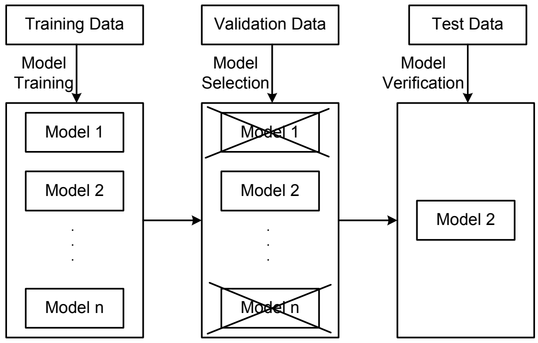

Figure 2.1: *Different data types used in the modeling process.*

In the identification process, several models are trained with training data. They are then selected with validation data. Finally, test data is used to verify the model quality.

2.4 Figures of Merit

The FOMs are used to quantify how well a model describes a system. Figure 2.2 shows how a FOM is obtained. The measured input signal is used to obtain the model output signal, which is compared with the measured output signal. The results are described by the FOM.

It can be defined in several ways, in time or in frequency domain, linear or logarithmic, parametric (a number) or nonparametric (graphic). Most commonly used FOMs will be described in this chapter.

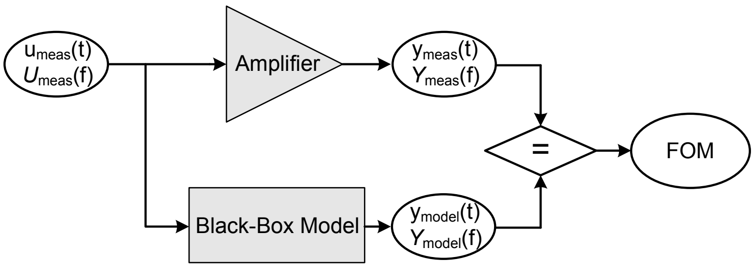

Figure 2.2: *Obtaining a FOM.*

2.4.1 Parametric FOMs

These FOMs are characterized by a number, summarizing the model performance.

Percent Variance Accounted For

The Percent Variance Accounted For (%VAF) is widely used in physiological system identification. It is a statistically based figure of merit, defined as [44]:

$$\%\text{VAF} = 100\left(1 - \frac{var(\mathbf{y} - \hat{\mathbf{y}})}{var(\mathbf{y})}\right) \tag{2.1}$$

Where $var(\cdot)$ is the variance, $\mathbf{y}$ and $\hat{\mathbf{y}}$ are the measured and model estimated output signal, respectively. It is a linear FOM, well suited for general modeling, and sensible to the synchronization between measurement and model signals. Very small differences in measured and modeled signals are almost imperceptible, due to the linear nature of this FOM.

Normalized Mean Square Error

The Normalized Mean Square Error (NMSE) is a frequently used metric to verify the model accuracy on a sample-per-sample basis [33, 36, 45–47]. It is defined as:

$$\text{NMSE} = 10\log\left\{\frac{\sum_{k=1}^{N}|y(k) - \hat{y}(k)|^2}{\sum_{k=1}^{N}|y(k)|^2}\right\} \tag{2.2}$$

The NMSE main characteristics are:

- Accounts for very small variations in signals;
- Well suited for measurement of the accuracy of the model in the presence of IMD, even when they are very small;
- Provides a logarithmic measurement.

Although it is apparently an efficient and parsimonious FOM, it reveals little information about the modeling of memory effects, nor which kind of memory effect was modeled.

The NMSE can be mathematically related to the raw EVM [48] as:

$$\text{EVM}_{\text{raw}} = 10^{\frac{\text{NMSE (dB)}}{20}} \tag{2.3}$$

Normalized Mean Square Error Problems

The main problems of the ordinary NMSE are its high sensibility to:

- Errors in linear components of the signal;
- Signal's synchronization.

Nonlinear distortions, which are really interesting from a nonlinear behavioral modeling perspective, have little weight in this metric.

Error in Linear Components

Fig. 2.3 shows two model output signals, and the respective NMSE related to the measured output signal. The black signal is obtained from a model without any modifications (NMSE = -34.2 dB). For the generation of the signal that has a NMSE of $-20.2\,\text{dB}$, a gain of 10% was applied only to the linear coefficients of the same model, thus introducing an intentional error in these coefficients. The FOM NMSE was strongly influenced by this 10% gain in linear coefficients, although no alterations were made in nonlinear coefficients of the model.

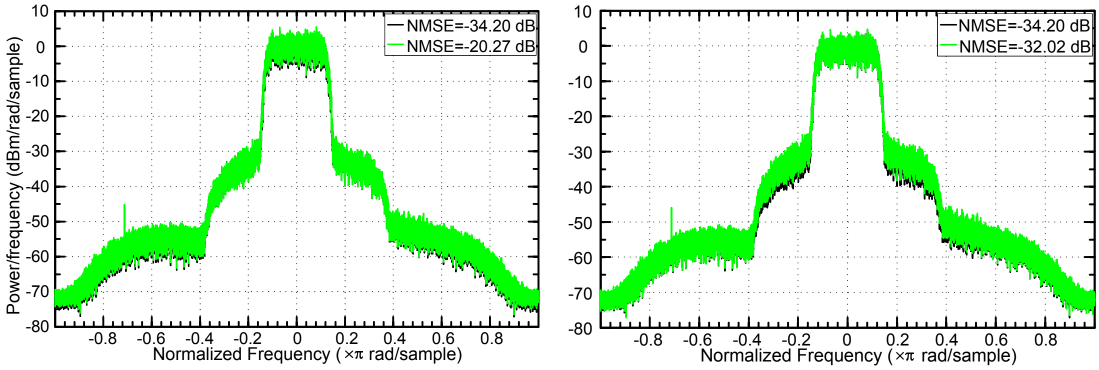

Figure 2.3: *Modified model output signals for NMSE performance analysis. The signal with* NMSE $= -20.2\,\text{dB}$ *was generated from a model with an increase of* 10% *at linear coefficients (left). The signal with* NMSE $= -32\,\text{dB}$ *was generated from a model with an increase of* 20% *at nonlinear coefficients (right).*

Doing a similar analysis, an increase of 20% was done only in nonlinear coefficients of the original model. An even higher error was expected, due to the higher error introduced in the nonlinear coefficients. Despite these alterations, the NMSE pointed out a mere $2\,\text{dB}$ difference, leading to the wrong conclusion that model with and without increase in nonlinear coefficients have a close performance. The results are also displayed in Fig. 2.3.

Synchronization Errors

Figure 2.4, shows two model output signals and their NMSE results relative to the measured output signal. The signal with a NMSE of $-1.5\,\text{dB}$ has a synchronization difference of 4 samples relative to the signal with NMSE of $-34.2\,\text{dB}$. If only the NMSE is used as a FOM, the model which produced an output signal with a NMSE of $-1.5\,\text{dB}$ would be badly classified, although the model output signals are almost the same, having only a synchronization error.

So, gain differences and synchronization errors have to be considered when applying NMSE to verify BM's accuracy.

Adjacent Channel Power Ratio

A method used to measure modeling performance and compare models in frequency domain is to evaluate the differences in the adjacent channel power ratio (ACPR) of the measured and modeled outputs and calculate its difference. This FOM is defined

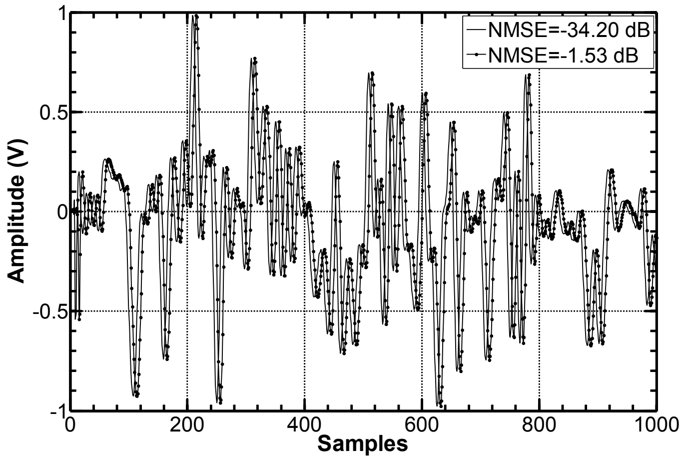

Figure 2.4: *Unsynchronized model output signals for NMSE performance analysis. The signal with a NMSE of* $-1.5\,\mathrm{dB}$ *has a synchronization difference of* 4 *samples relative to the signal with NMSE of* $-34.2\,\mathrm{dB}$.

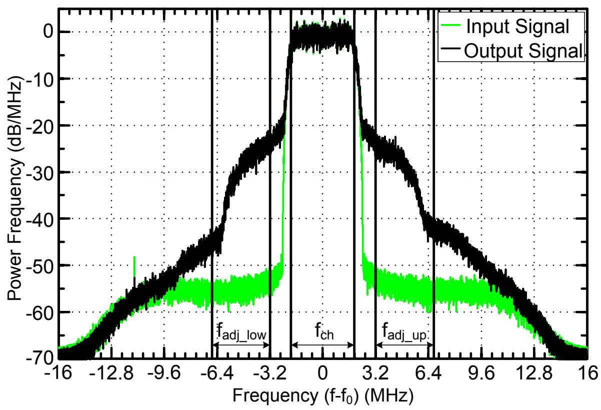

Figure 2.5: *WCDMA signal and ranges used for ACPR calculation.*

as:

$$\begin{aligned}\Delta\mathrm{ACPR} =& \mathrm{ACPR}^{\mathrm{meas}} - \mathrm{ACPR}^{\mathrm{mod}} \\ \mathrm{ACPR} =& 10\log\left\{\frac{\int_{f_{\mathrm{adj}}} |Y(f)|^2 df}{\int_{f_{\mathrm{ch}}} |Y(f)|^2 df}\right\}\end{aligned} \tag{2.4}$$

here $Y(f)$ is the Fourier transform of the corresponding signal, f_{adj} the frequency band of the adjacent channel, and f_{ch} the carrier frequency range. The output filter included in a similar FOM calculation (adjacent channel leakage ratio – ACLR) according to standard [48] was neglected in this case. Figure 2.5 shows the input and output signal spectra for a WCDMA signal and the respective frequency ranges used for the ACPR calculation.

This is a very useful FOM, since it is also widely employed in the frequency domain for amplifiers characterization, when amplifier output spectra are tested with a spectral mask defined by a specific standard.

The same analysis done before in section 2.4.1, increasing linear coefficients by 10% and nonlinear coefficients by 20%, was repeated for ΔACPR, and the results are listed in Table 2.1. The results were not so conclusive, close to the error free ΔACPR value, and bring no information about which kind of inaccuracy the model has (e. g., bad fitting of linear or nonlinear memory). Nevertheless, the difference between linear and nonlinear errors is smaller.

Table 2.1: *Error in parameters analysis for ΔACPR*

Model	0.18
10% error linear coef.	−0.58
20% error nonlinear coef.	1.65

2.4.2 Nonparametric FOMs

Although parametric FOMs offer a very parsimonious way to quantify a model, they are not optimal to avail how much memory of the system is considered by the obtained model. Graphical FOMs may deliver better results in this case.

AM/AM – AM/PM curves

Boundary lines of the AM/AM (gain compression) – AM/PM (phase distortion) conversion plots are an indication of memory effects. Figure 2.6 shows an example, for a WCDMA signal applied to a PA operating at 5 dB input back off. The points spread clearly indicate that the amplifier presents linear and nonlinear memory.

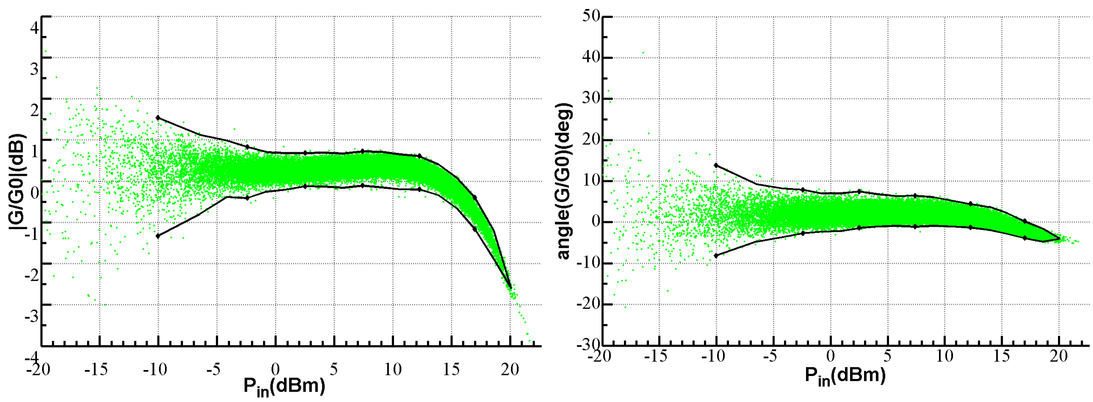

Figure 2.6: *AM/AM – AM/PM conversion boundary lines of a WCDMA signal applied to a PA.*

By the use of boundary lines instead of the full conversion plot it is also possible to compare several AM/AM conversion plots within the same graph. As an example, Fig. 2.7 shows the boundary lines of the AM/AM conversion of the reference and the modeled output signals for three different models. The best performance was achieved by Model3, with the closest approach to the reference lines.

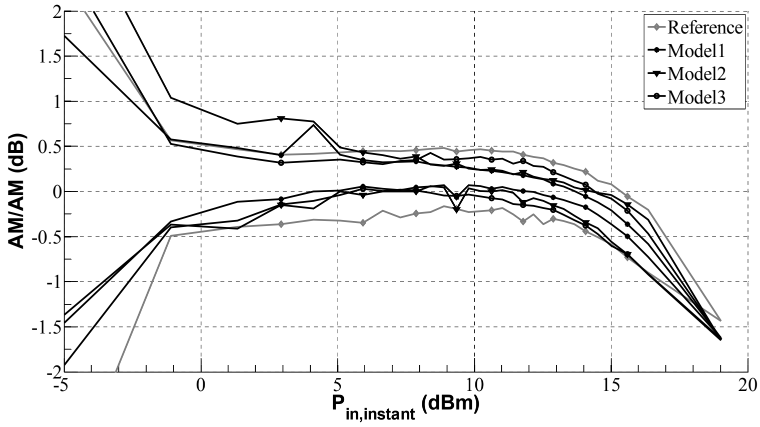

Figure 2.7: *AM/AM-conversion boundary lines of the reference and several modeled output signals.*

Another possibility is to visualize the AM/AM and AM/PM signal characteristics simultaneously, in a 3-D plot. Figure 2.8 shows a mean curve in black that denotes the nonlinearity of an amplifier output signal. The other curves are projections of this curve in the xy, xz, and yz planes.

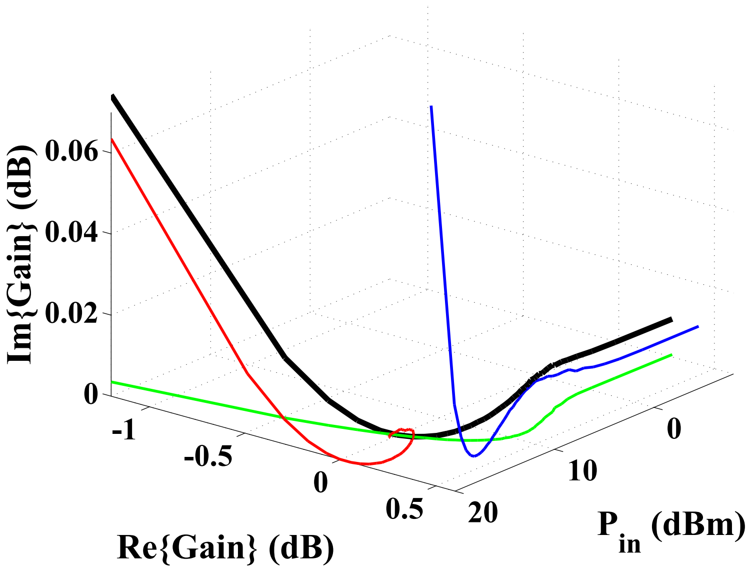

Figure 2.8: *AM/AM – AM/PM mean curve denotes the amplifier static nonlinear behavior.*

Extending this concept to include also memory effects, it is possible to draw a 3-D graph with polygons surrounding this line (see Fig. 2.9). They stand for memory outer borders in 3-D, as boundary lines in 2-D in Fig. 2.6.

This figure shows that an amplifier BM should take into consideration AM/AM and AM/PM variations in order to be complete. Also a correct reproduction of PA memory is important, since it is essential for accurate adaptive pre-distortion purposes [5].

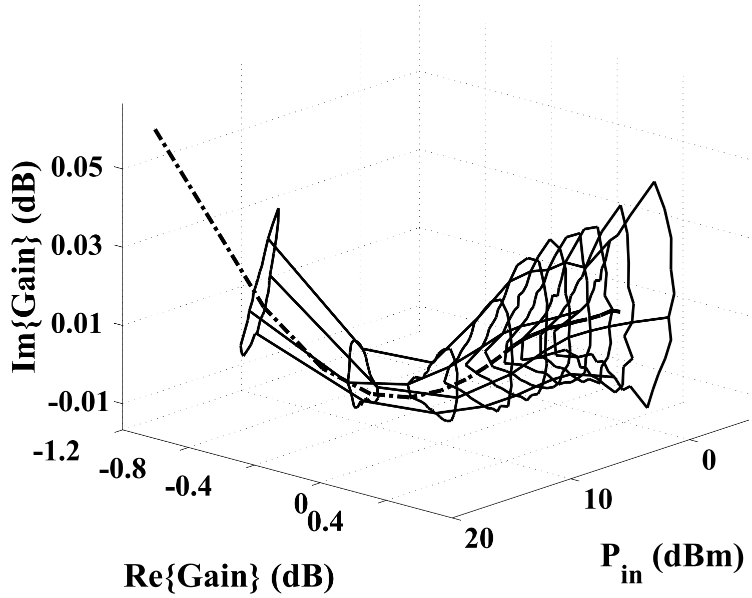

Figure 2.9: *3-D representation of the signal AM/AM and AM/PM variations.*

Complimentary Cumulative Density Function

Another useful indication of the quality of a model's output signal is a comparison of the Complimentary Cumulative Density Function (CCDF) of measured and modeled outputs. Figure 2.10 displays an example of this procedure for three given models. Throughout this FOM, the comparison of model's capability to reproduce peaks encountered in the output signal is possible. For this particular case, the performance of the models is very close, and other FOMs are needed to decide which model is more accurate.

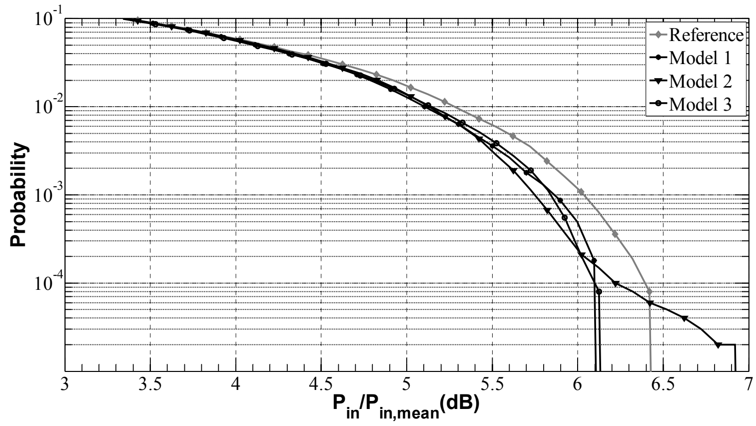

Figure 2.10: *CCDF of the reference and several modeled output signals.*

Coherence

The coherence is a function of the power spectral density (PSD) and the cross PSD of the input/output signal. It is defined as [49]:

$$C_{uy}(f) = \frac{|S_{uy}(f)|^2}{S_{uu}(f)S_{yy}(f)} \tag{2.5}$$

Where $S_{uu}(f)$ and $S_{yy}(f)$ are the input/output PSD, respectively, and $S_{uy}(f)$ is the cross PSD, defined as

$$S_{\mathrm{uy}}(f) = \langle \mathbf{u}(f)\mathbf{y}^*(f)\rangle \tag{2.6}$$

with $\langle \mathbf{u}(f)\rangle = 0$ and $\langle \mathbf{y}(f)\rangle = 0$, where $< \cdot >$ is the expectation operator. The auto PSD is derived in a similar way. If the coherence is estimated between two linearly related signals, the output for all frequencies is $C_{uy}(f) = 1$.

Figure 2.11 shows the coherence graphs between two models and measured output signal: a linear (a FIR filter) and a nonlinear dynamic model (Wiener model). This graph is a measure for model's representation of nonlinear effects [23], or if there are any unmodeled linear dynamics in the model. If there are unconsidered dynamics in the model output signal, the value tends to zero at that frequency.

The improvements achieved by the nonlinear model are clear in Fig. 2.11, once there are less values near zero for the nonlinear curve than in the linear one. No information about dynamics can be extracted using only this parametric FOM.

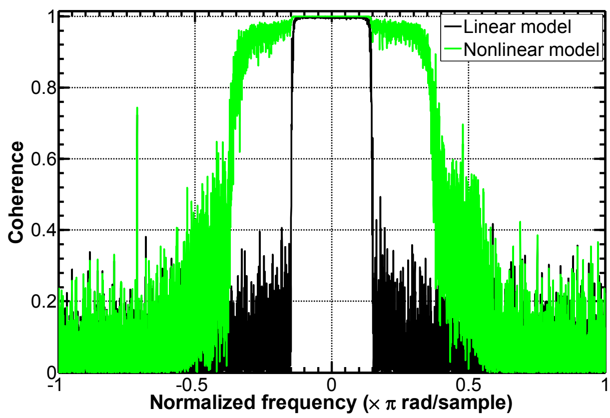

Figure 2.11: *Coherence between a linear (FIR filter), a nonlinear model (Wiener), and the measured output signal of a given amplifier.*

2.5 Conclusion

This chapter depicts issues about system modeling. In excitation signal design, it was shown that the Gaussian noise is the most suitable signal, but unrealizable in practice. The stimulus signals used for modeling are in general the ones also used in normal operating conditions by the system. It was also explained how and why data is partitioned in modeling, validation and test data in the estimation process. Finally, parametric and nonparametric FOMs were displayed. The former is frequently used in estimation algorithms due to its easiness of evaluation, and the latter is more suitable to the analysis of system dynamics represented by models under test.

Chapter 3

Linear Estimation and Linear Systems Estimation

3.1 Introduction

The dynamic part of an amplifier model represents memory effects, constituted by filters. Its coefficients can be estimated together with the nonlinear model part (as in Volterra series based models), or separately (as for block models), using a linear estimation technique. Several methods can be used for the linear estimation of linear models in time or frequency domain. The main properties of linear estimators are [40]:

- A unique optimum exists;
- A "one shot" solution can be computed analytically;
- Many numerically stable and fast algorithms are published (LS, TLS, LMS, RLS etc.).

In this chapter, a linear estimation overview for linear systems will be given, and different strategies will be commented.

3.2 Least Squares Method

The Least Squares (LS) is the most commonly used solution for linear optimization problems. Least square sense means the minimal sum of squared error loss function values, or the best linear combination of regressors. A detailed LS algorithm derivation can be found in [50]. All equations here described are for single input–single output systems.

The LS equation is defined as:

$$\hat{\mathbf{h}} = (\mathbf{U}^{\mathrm{H}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{y} \tag{3.1}$$

where $\hat{\mathbf{h}}$ is the estimated parameter vector, $\mathbf{U}$ is the regression matrix, $\mathbf{U}^{\mathrm{H}}\mathbf{U}$ is the Hessian, and $\mathbf{y}$ the output vector.

A problem when inverting very large Hessians from (3.1) is the insufficient memory. In this case, it is recommended to use the Strassen inversion method [51]:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\boldsymbol{\Sigma}^{-1} \\ -\boldsymbol{\Sigma}^{-1}\mathbf{C}\mathbf{A}^{-1} & \boldsymbol{\Sigma}^{-1} \end{bmatrix} \tag{3.2}$$

where $\mathbf{A}$ and its Schur complement $\boldsymbol{\Sigma} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})$ are the only terms requiring inversion.

An extension of the LS is the weighted LS:

$$\hat{\mathbf{h}} = (\mathbf{U}^{\mathrm{H}}\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{H}}\mathbf{W}\mathbf{y} \tag{3.3}$$

where the weight matrix $\mathbf{W}$ is composed as:

$$\mathbf{W} = \begin{bmatrix} W_{11} & 0 & \dots & 0 \\ 0 & W_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_{NN} \end{bmatrix} \tag{3.4}$$

where N is the number of samples of the input observation signal. Using this weighted estimator it is possible to emphasize the contributions of some sets of modeling data, correcting the ordinary LS method, where the contribution of all data is assumed to be equal. A drawback is that this method becomes inefficient for very large data-sets, which results in very large matrix $\mathbf{W}$.

The way that the LS regression matrix is composed allows the solution of a wide range of linear estimation problems.

Of special interest for amplifier modeling is the LS for FIR estimation, since this filter is the linear block of some modular models (Wiener, Hammerstein, Wiener-Hammerstein and the respective cascades) [23].

3.2.1 LS for FIR Filter Estimation

The FIR filter components are unit-delays, multipliers and adders. The order of the filter is given by M in Fig. 3.1. The output $y(k)$ is the convolution of the input $u(k)$ with the filter impulse response function (IRF) vector $\hat{\mathbf{h}}$.

The FIR regression matrix $\mathbf{U}$ for N values, a memory length M and the parameter vector $\hat{\mathbf{h}}$ is:

$$\hat{\mathbf{h}} = \begin{bmatrix} h_0 & h_1 & \dots & h_M \end{bmatrix}^T \tag{3.5}$$

$$\mathbf{U} = \begin{bmatrix} u(M+1) & u(M) & \dots & u(1) \\ u(M+2) & u(M+1) & \dots & u(2) \\ \vdots & \vdots & \ddots & \vdots \\ u(N) & u(N-1) & \dots & u(N-M) \end{bmatrix} \tag{3.6}$$

For accurate estimations, N should be at least 10 times bigger than M (empirical observation).

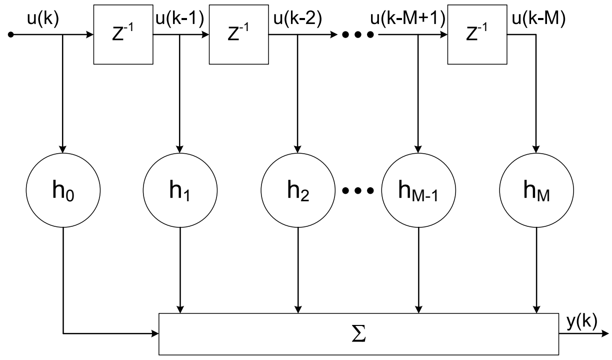

Figure 3.1: *Schematic of a FIR filter.*

3.2.2 Correlation Methods

Another way to represent the LS estimator (3.1) is through correlations. Formulating the general regression matrix as:

$$\mathbf{U}^{\mathrm{H}} = \begin{bmatrix} u_1^*(1) & u_1^*(2) & \dots & u_1^*(N) \\ u_2^*(1) & u_2^*(2) & \dots & u_2^*(N) \\ \vdots & \vdots & \ddots & \vdots \\ u_n^*(1) & u_n^*(2) & \dots & u_n^*(N) \end{bmatrix} \tag{3.7}$$

and:

$$\mathbf{U} = \begin{bmatrix} u_1(1) & u_2(1) & \dots & u_n(1) \\ u_1(2) & u_2(2) & \dots & u_n(2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(N) & u_2(N) & \dots & u_n(N) \end{bmatrix} \tag{3.8}$$

The Hessian $\mathbf{H} = \mathbf{U}^{\mathrm{H}}\mathbf{U}$ in (3.1) can be written as:

$$\mathbf{H} = \begin{bmatrix} \sum_{i=1}^{N} |\mathbf{u}_1(i)|^2 & \sum_{i=1}^{N} \mathbf{u}_1^*(i)\mathbf{u}_2(i) & \dots & \sum_{i=1}^{N} \mathbf{u}_1^*(i)\mathbf{u}_n(i) \\ \sum_{i=1}^{N} \mathbf{u}_2^*(i)\mathbf{u}_1(i) & \sum_{i=1}^{N} |\mathbf{u}_2(i)|^2 & \dots & \sum_{i=1}^{N} \mathbf{u}_2^*(i)\mathbf{u}_n(i) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} \mathbf{u}_n^*(i)\mathbf{u}_1(i) & \sum_{i=1}^{N} \mathbf{u}_n^*(i)\mathbf{u}_2(i) & \dots & \sum_{i=1}^{N} |\mathbf{u}_n(i)|^2 \end{bmatrix} \tag{3.9}$$

Which is the autocorrelation matrix $\mathbf{R}_{\mathrm{uu}}$. Using a similar procedure, $\mathbf{U}^{\mathrm{H}}\mathbf{y}$ in (3.1) is:

$$\mathbf{U}^{\mathrm{H}}\mathbf{y} = \begin{bmatrix} \sum_{i=1}^{N} \mathbf{u}_1^*(i)\mathbf{y}_1(i) \\ \sum_{i=1}^{N} \mathbf{u}_2^*(i)\mathbf{y}_1(i) \\ \vdots \\ \sum_{i=1}^{N} \mathbf{u}_n^*(i)\mathbf{y}_1(i) \end{bmatrix} \tag{3.10}$$

This expression is the input/output signal cross-correlation $\mathbf{r}_{\mathrm{uy}}$. The resulting LS solution is [40]:

$$\hat{\mathbf{h}} = \hat{\mathbf{R}}_{\mathrm{uu}}^{-1}\hat{\mathbf{r}}_{\mathrm{uy}} \tag{3.11}$$

The main difficulty is that neither $\hat{\mathbf{R}}_{\mathrm{uu}}$ nor $\hat{\mathbf{r}}_{\mathrm{uy}}$ are available in advance and have to be estimated. Equation (3.11) is also known as Wiener filter and can be used for adaptive filter techniques (see [52]).

This is a good approximation only if very long training sequences are used for the estimation of the autocorrelation function.

3.2.3 Singular Value Decomposition Techniques (SVD)

If the system input signal is a band-limited white Gaussian noise, an alternative way to find a suitable form for the LS estimator of the IRF $\mathbf{h}$ is to apply the SVD, due to its de-noising possibility [39].

Band-limited white Gaussian noise are signal conditions of some modern standard communication signals, like WCDMA and Wimax. This technique is computationally efficient and very suitable if signal-to-noise ratio (SNR) is low [53]. The model assumed is in Fig. 3.2.

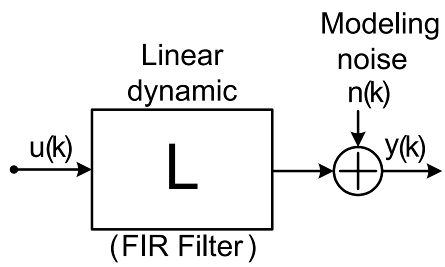

Figure 3.2: *Model assumed for a linear filter estimation.*

The pseudo-inverse derivation [12, 53] for complex data is:

$$\begin{aligned} \hat{\mathbf{h}} &= \mathbf{R}_{\mathrm{uu}}^{-1}\mathbf{r}_{\mathrm{uy}} + \hat{\mathbf{R}}_{\mathrm{uu}}^{-1}\hat{\mathbf{r}}_{\mathrm{un}} && (3.12) \\ &= \mathbf{h} + \hat{\mathbf{R}}_{\mathrm{uu}}^{-1}\hat{\mathbf{r}}_{\mathrm{un}} && (3.13) \\ &= \mathbf{V}\mathbf{S}^{-1}\mathbf{V}^{\mathrm{H}}\mathbf{V}\mathbf{S}\mathbf{V}^{\mathrm{H}}h + \mathbf{V}\mathbf{S}^{-1}\mathbf{V}^{\mathrm{H}}\hat{\mathbf{r}}_{\mathrm{un}} && (3.14) \end{aligned}$$

with $\mathbf{r}_{\mathrm{uy}} = \mathbf{R}_{\mathrm{uu}}\mathbf{h}$ and $\mathbf{R}_{\mathrm{uu}}^{-1} = \mathbf{V}\mathbf{S}^{-1}\mathbf{V}^{\mathrm{H}}$. The vector $\hat{\mathbf{r}}_{\mathrm{un}}$ is the input/output noise correlation, $\mathbf{V}$ is a matrix composed of the singular vectors and $\mathbf{S}$ is a diagonal matrix formed by the singular values. $\hat{\mathbf{R}}_{\mathrm{uu}}$ are the estimated input correlation matrices. Substituting $\mathbf{V}^{\mathrm{H}}\mathbf{h} = \varsigma$ and $\mathbf{V}^{\mathrm{H}}\hat{\mathbf{r}}_{\mathrm{un}} = \eta$ yields to:

$$\begin{aligned} \hat{\mathbf{h}} &= \mathbf{V}\mathbf{V}^{\mathrm{H}}\mathbf{h} + \mathbf{V}\mathbf{S}^{-1}\mathbf{V}^{\mathrm{H}}\hat{\mathbf{r}}_{\mathrm{un}} && (3.15) \\ &= \sum_{j=1}^{T}\left(\varsigma_j + \frac{\eta_j}{s_j}\right)\mathbf{v}_j && (3.16) \end{aligned}$$

where $\mathbf{s}_j$ are the singular values. Only terms for which

$$|\varsigma_j| \geq \frac{|\eta_j|}{s_j} \tag{3.17}$$

contribute to the estimator in (3.16). The Minimum Description Length (MDL) cost function is defined as [12]:

$$\mathrm{MDL}(n_{\mathrm{sv}}) = \left(1 + n_{\mathrm{sv}}\frac{\log(N)}{N}\right)\sum_{k=1}^{N}(y(k) - \hat{y}(k))^2 \tag{3.18}$$

where n_{sv} is the number of singular vectors, N is the total number of input/output realizations, $y(k)$ is the reference output and $\hat{y}(k)$ is the model's output. Using (3.17) and (3.18), it is possible to separate the necessary singular vectors $\mathbf{s}$ for the estimator.

Another possibility is to develop an algorithm considering the noise-free output variance of a linear system without bias at its input [23]:

$$\begin{aligned}
\frac{1}{N}\sum_{k=1}^{N} y^2(k) &= \frac{1}{N}\sum_{k=1}^{N}\left(\sum_{\tau_1=0}^{M-1} h(\tau_1)u(k-\tau_1)\right)\left(\sum_{\tau_2=0}^{M-1} h(\tau_2)u(k-\tau_2)\right) && (3.19)\\
&= \sum_{\tau_1=0}^{M-1}\sum_{\tau_2=0}^{M-1} h(\tau_1)h(\tau_2)\frac{1}{N}\sum_{k=1}^{N} u(k-\tau_1)u(k-\tau_2) && (3.20)\\
&= \mathbf{h}^{\mathrm{H}}\mathbf{R}_{uu}\mathbf{h} && (3.21)\\
&= \mathbf{h}^{\mathrm{H}}\mathbf{V}\mathbf{S}\mathbf{V}^{\mathrm{H}}\mathbf{h} && (3.22)
\end{aligned}$$

A rearrangement of the taps can be carried out considering the contribution of each component γ_i that composes the output variance:

$$\begin{aligned}
\gamma_i &= \sum_{i=1}^{M} s_i\left(\mathbf{v}_i^{\mathrm{H}}\hat{\mathbf{h}}\right)\left(\mathbf{v}_i^{\mathrm{H}}\hat{\mathbf{h}}\right)^{\mathrm{H}} && (3.23)\\
&= \frac{1}{s_i}\mathrm{diag}\left[\left(\mathbf{V}^{\mathrm{H}}\mathbf{r}_{\mathrm{uz}}\right)\left(\mathbf{V}^{\mathrm{H}}\mathbf{r}_{\mathrm{uz}}\right)^{\mathrm{H}}\right] && (3.24)
\end{aligned}$$

Where $\mathrm{diag}(\cdot)$ is the main diagonal of a matrix. Then, after estimating $\hat{\mathbf{R}}_{\mathrm{uu}}$ and $\hat{\mathbf{r}}_{\mathrm{uy}}$, equation (3.23) is calculated by means of the SVD of $\hat{\mathbf{R}}_{\mathrm{uu}}$, and the results can be sorted in decreasing order. This is the final IRF estimate.

Benefits of SVD techniques

The rearranging of the taps brings some benefits for the linear estimation. Fig. 3.3 shows a comparison of the estimated output signals using ordinary FIR LS and SVD techniques. An improvement can be seen in out-of-band components, where results are closer to the measured signal. Also in Fig. 3.3, the resulting taps of these estimators are displayed. If the filter memory estimation (the value that the IRF magnitude is zero) was done by the ordinary LS estimator, the results would be around 10 taps. Using SVD techniques, the first tap with a value around zero is observed at position 5, so the filter function has more energy concentrated in less taps, allowing the use of models with less parameters and with the same efficiency.

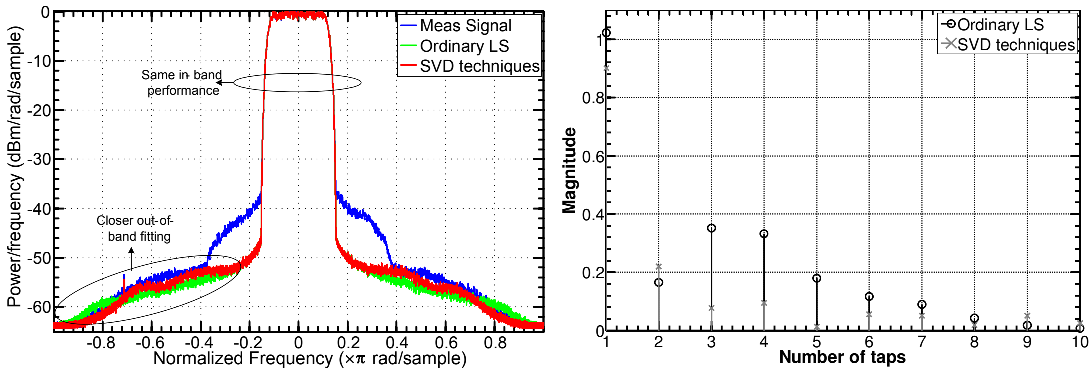

Figure 3.3: *Spectral results for the measured and modeled signals using ordinary FIR LS and SVD techniques. At the right side respective taps magnitudes are displayed.*

3.3 Total Least Squares

The LS estimation technique is based on minimizing an error given by:

$$e_{LS} = \min ||\mathbf{y} - \hat{\mathbf{y}}||_2 \tag{3.25}$$

and once the minimum is found, any h satisfying:

$$\mathbf{Uh} = \mathbf{y} \tag{3.26}$$

is the LS solution.

The Total Least Squares (TLS) [39, 54] was developed to account for errors in the LS regression matrix $\mathbf{U}$ also, and not only in the output vector. The TLS problem is formulated as minimizing the error given by:

$$e_{\text{TLS}} = \min ||[\mathbf{U}|\mathbf{y}] - [\hat{\mathbf{U}}|\hat{\mathbf{y}}]||_F \tag{3.27}$$

Where $[\mathbf{U}|\mathbf{y}]$ indicates that $\mathbf{U}$ and $\mathbf{y}$ are stacked side-by-side, and $||\cdot||_F$ is the Frobenius norm. The TLS takes input/output distortions into account, a situation that occurs in practice.

The difference between these two methods is displayed in Fig. 3.4 for the one parameter equation $\mathbf{y} = h\mathbf{u}$. The LS approach minimizes the vertical distance to line, and the TLS minimizes the total distance to line [39].

The entire solution is obtained by using SVD, and is explained in details in [39, 54]. Let $C = [\mathbf{U}|\mathbf{y}]$, then:

$$C = \mathbf{U}diag(s_1, s_2, \ldots, s_{n+1})V^{\text{H}} = \sum_{k=1}^{n+1} s_k \mathbf{u}_k \mathbf{v}_k^{\text{H}} \tag{3.28}$$

with $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n+1}]$. The final solution is:

$$\mathbf{h} = -\frac{\mathbf{v}_{n+1}(1 \ldots n)}{\mathbf{v}_{n+1}(n+1)} \tag{3.29}$$

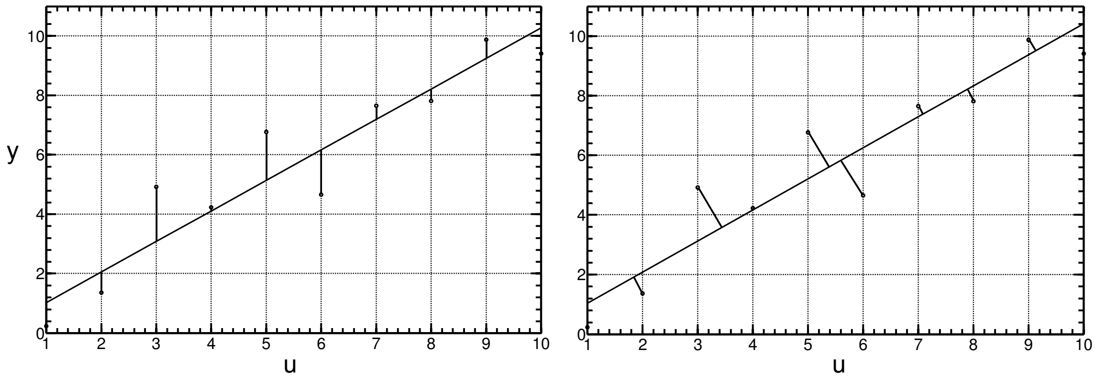

Figure 3.4: *LS (left) and TLS (right) solution for the one parameter equation.*

3.4 Frequency-Domain Estimation

Using the input power spectrum and the input/output cross-power spectrum it is possible to estimate the frequency response:

$$\hat{H}(f) = \frac{S_{\mathrm{uy}}(f)}{S_{\mathrm{uu}}(f)} \tag{3.30}$$

where $\hat{H}(f)$ is the estimated frequency response, $S_{\mathrm{uy}}(f)$ the input/output cross-power spectrum and $S_{\mathrm{uu}}(f)$ the input power spectrum.

$S_{\mathrm{uy}}(f)$ is defined as the Fourier transform of the cross-correlation function $\mathbf{r}_{\mathrm{uy}}(\tau)$

$$S_{\mathrm{uy}}(f) = \sum_{-\infty}^{+\infty} \mathbf{r}_{\mathrm{uy}}(\tau) e^{-j2\pi f\tau} d\tau \tag{3.31}$$

Other properties and definitions can be found in [55]. $S_{\mathrm{uu}}(f)$ can be derived in a similar way.

This method is called the indirect method of power spectrum estimation [56].

The direct method is calculating the expectation of the Fourier transform of the input/output signal.

The main disadvantage of this approach is that much averaging is necessary to reduce the random error to acceptable levels when signals are noisy.

3.5 Conclusion

Different types of estimators can be used to determine the dynamic part of a BM, or even to estimate the entire BM, if the BM structure is "linear in parameters".

The techniques presented in this chapter are less computational intensive than non-linear optimization techniques, and are a fast way to calculate the solution.

Due to these characteristics, linear estimation techniques have great potential to be widely used and implemented in hardware.

Chapter 4

Nonlinear Systems Estimation

4.1 Introduction

A RF PA is a typical nonlinear system. Even when the transistor is operating in the linear region, driven with small variance input signals, the output signal has nonlinear components, due to the physics of the transistor.

Different methods can be used for nonlinear systems estimation in the time or in the frequency domain. Nonlinear estimation techniques are more powerful compared to linear ones, but they have some undesirable properties [40]:

- Multiple minima in the cost function;
- Iterative solutions are often necessary;
- Numerically stable and efficient algorithms are still under research;
- Nonlinear estimation algorithms can present better results for estimation, but the convergence rate and the initial parameters have to be well chosen.

Due to these characteristics, and also because models used in this work are linear in parameters, similar algorithms as for the linear systems estimation will be used.

4.1.1 Classification of Nonlinear Systems

In [25] a complete classification of nonlinear systems is given. If any of the following phenomena occurs, a nonlinear dynamic model has to be used:

1. Asymmetric responses to symmetric input signal changes (ASYM);
2. Generation of higher-order harmonics in response to a sinusoidal input (HARM);
3. Input multiplicity, means that one steady-state response corresponds to more than one steady-state input (IM);
4. Output multiplicity, means that one steady-state input corresponds to more than one steady-state output (OM);
5. Generation of sub-harmonics in response to any periodic input (SHAM);

6. Highly irregular responses to simple inputs like impulses, steps or sinusoids (CHAOS);
7. Input-dependent stability (IDS).

A nonlinear system is classified due to phenomena presence as follows:

- Mild: ASYM, HARM, IM;
- Intermediate: IDS;
- Strong: OM, SHAM, CHAOS.

A PA BM remains in the mildly nonlinear class. The known PAs to be modeled present these characteristics in normal operation conditions, when tested with a sinusoid stimuli. None of the other phenomena (OM, SHAM, CHAOS or IDS), that imply the need of intermediate or strong nonlinear dynamic models, were observed in amplifier measurements.

Classification of Power Amplifier Behavioral Models

The classification of PA BMs used in this work is as in [21]:

- Memoryless: The output envelope reacts instantaneously to variations in the input envelope;
- Linear memory: BM that account for envelope memory effects attributable to the input and output matching networks' frequency characteristics;
- Nonlinear memory: Dynamic interaction of nonlinearities through a dynamic network.

Figure 4.1 is used by the authors to classify the various BMs. Memoryless models are represented by the block "Nonlinear/Memoryless". Linear memory models are models that accounts for the "Nonlinear/Memoryless", H(ω) and O(ω) blocks (matching networks). Models that care for nonlinear memory contains all previous mentioned blocks, and the feedback path with the block F(ω), attributed to electrothermal and/or bias circuitry dynamics.

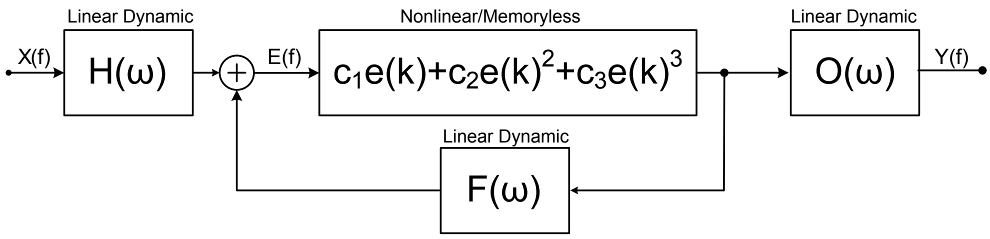

Figure 4.1: *A PA representation using a nonlinear feedback structure.*

4.1.2 Volterra Series

Volterra series (VS) account for mildly nonlinear class of nonlinear systems and has the property of dynamic interaction of nonlinearities, so it is well suited for the description of PA.

The finite, discrete VS model is given by [18]:

$$y_V(k) = \sum_{p=0}^{P} \sum_{\tau_1=0}^{M-1} \cdots \sum_{\tau_p=0}^{M-1} h_p(\tau_1, \cdots, \tau_p) u(k-\tau_1) \dots u(k-\tau_p) \tag{4.1}$$

Where h_p is the kernel of order P, k and τ are discrete indices of the sampling interval, and M is the memory length. The sampling interval must be selected to cover the needed input/output signal's bandwidth.

The main disadvantage of a VS based BM is the number of parameters necessary to estimate, and consequently, to represent the model. A Volterra model using 5 delay taps needs 5, 125, 625, and 3125 parameters for the 1^{st}, 3^{rd}, 5^{th}, and 7^{th} order kernels, respectively. These values are not practical, since an estimation using so many coefficients is very computational intensive, even for actual computers.

By using the symmetry condition, the complexity of the Volterra kernels as a function of the order of nonlinearity is given by the binomial [35]:

$$\binom{M+p-1}{p} \tag{4.2}$$

where M is the number of delay taps used and p is the order of the kernel. Using (4.2), the above cited model is reduced to 5, 45, 126, and 330 parameters. Unfortunately, this equation is valid only for real valued signals.

Complex Valued Baseband Volterra series

In order to obtain the best model performance, it is necessary to adapt the BMs under study to the modern PA input/output industry standard signals, once these models are designed for linearization purposes. The excitation signals are complex valued, and as a practical issue only first-zone filtered (baseband) equivalent BMs are frequently used, due to the difficulties to implement bandpass models in hardware.

The model complexity grows significantly for baseband VS using complex signals, represented until 7^{th} order in (4.3) for symmetric kernels.

$$
\begin{aligned}
y(k)=&\sum_{i=0}^{M-1} h_1(i)u(k-i)+\sum_{i=j}^{M-1}\sum_{j=0}^{M-1}\sum_{l=0}^{M-1} h_3(i,j,l)u(k-i)u(k-j)u^*(k-l)+\\
&+\sum_{i=j}^{M-1}\sum_{j=l}^{M-1}\sum_{l=0}^{M-1}\sum_{m=n}^{M-1}\sum_{n=0}^{M-1} h_5(i,j,l,m,n)u(k-i)u(k-j)u(k-l)u^*(k-m)u^*(k-n)+\\
&+\sum_{i=j}^{M-1}\sum_{j=l}^{M-1}\sum_{l=m}^{M-1}\sum_{m=0}^{M-1}\sum_{n=r}^{M-1}\sum_{r=s}^{M-1}\sum_{s=0}^{M-1} h_7(i,j,l,m,n,r,s).\\
&.\ u(k-i)u(k-j)u(k-l)u(k-m)u^*(k-n)u^*(k-r)u^*(k-s)
\end{aligned} \tag{4.3}
$$

A closed form for determining the number of independent terms for baseband VS using complex signals is the binomial:

$$
\binom{M+\lfloor p/2\rfloor}{\lfloor p/2\rfloor}\binom{M+\lfloor p/2\rfloor}{\lfloor p/2\rfloor+1}\frac{1}{M+\lfloor p/2\rfloor} \tag{4.4}
$$

Where $\lfloor\cdot\rfloor$ is the floor operation.

As an example, the numbers of parameters for a system using 4 delay taps are 4, 40, 200, and 700 for the 1^{st}, 3^{rd}, 5^{th}, and 7^{th} order symmetric kernels. The use of the Volterra kernel symmetry property is necessary in the model extraction process, since it eliminates the linear dependent columns of the kernel to be estimated.

Several techniques are employed to estimate VS. If the system is memoryless, VS are reduced to a Taylor series and can be estimated as described in section 4.2. If the system has only linear memory, it can be estimated using the techniques listed in section 4.3. If the system presents only nonlinear memory or linear and nonlinear memory, some strategies described in section 4.4 can be employed.

Model Memory Estimation

The memory estimation is crucial since it determines the number of delays necessary in a VS and has direct influence onto the model performance. The system memory length can be estimated observing when the value of the transfer function of the first order Volterra kernel decays near zero [24].

Minimum Memory Length

The minimum number of lags along each dimension of the Volterra kernel required to represent it in discrete time domain is given by [24]:

$$
\mathrm{L_{min}} = 2\mathrm{B_s}\mu \tag{4.5}
$$

Where $\mathrm{B_s}$ (Hz) denotes the baseband bandwidth of the measured output signal and μ is the effective kernel memory or the correlation time over which the kernel has significant values.

Maximum Memory Length

A limit for the kernel maximum memory length is where it shows no exponential decaying, but some oscillations around zero [24] that can be considered modeling noise. This point limits the number of lags used for the next estimation. In [14], it was determined by squaring the kernels magnitude and calculating where they concentrate approximately 90% of the total kernel energy. An example of this procedure is shown in Fig. 4.2, where an arrow shows where the kernel was limited.

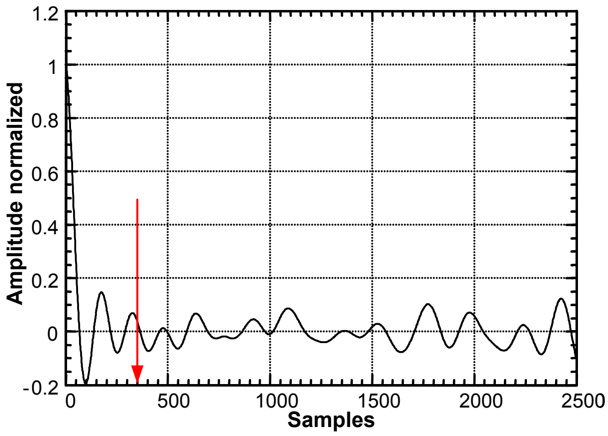

Figure 4.2: *Example of a first order kernel estimation. The arrow shows where the memory was limited for the next estimation.*

Minimum and maximum memory length are important information obtained from data, since no prior knowledge about the amplifier's memory are known in advance.

The kernel is limited to the smallest possible memory length for practical applications, in order to generate parsimonious models.

4.2 Nonlinear Memoryless

The nonlinear part of an amplifier model represents the IMD, or the static part, and is usually composed of polynomials or other nonlinear functions (e. g., tangent-sigmoids, look-up tables). These models do not account for dynamics of the system.

In this section, the memoryless nonlinear estimation based on measurements will be covered, and different strategies will be commented.

4.2.1 Power Series

A nonlinear system can be represented by a power series:

$$y(k) = \sum_{p=0}^{P} c_p u^p(k) \tag{4.6}$$

where c_p are the polynomial coefficients and P is the order.

Polynomial LS Estimation

A simple form to estimate a power series is using linear regression methods, as polynomial coefficients are linear in parameters. The polynomial regression matrix $\mathbf{U}$ for N measurements, a polynomial degree P and the parameter vector $\hat{\boldsymbol{\theta}}_P$ is [40]:

$$\mathbf{U} = \begin{bmatrix} 1 & u(1) & u^2(1) & \dots & u^P(1) \\ 1 & u(2) & u^2(2) & \dots & u^P(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u(N) & u^2(N) & \dots & u^P(N) \end{bmatrix} \tag{4.7}$$

$$\hat{\boldsymbol{\theta}}_P = \begin{bmatrix} c_0 & c_1 & \dots & c_P \end{bmatrix}^T \tag{4.8}$$

Then (3.1) can be applied. This regression matrix results in a Hessian with a high condition number (CN), defined as the ratio of the largest to smallest singular value in the singular value decomposition of a matrix [57]. A large CN is not desirable in the estimation process, as it implies that small errors in the input can cause large errors in the output.

The use of orthogonal polynomials can improve the Hessian CN for input signals that these polynomials were derived. The regressors are closer to the ideal situation for a Hessian (regressors mutually orthogonal).

For real valued input signals, Chebyshev (derived for single tones) and Hermite (derived for Gaussian distribution) polynomials are typically applied. For complex Gaussian baseband input signals, a derivation is found in [28].

Chebyshev Polynomials

The Chebyshev polynomials basis function is bounded between the interval $[-1, 1]$ for inputs between $[-1, 1]$, and this leads to similar regressors variances. The model input signals should be normalized to this input range. The modeling error near ± 1 is heavily weighted in this kind of polynomial [18]. This behavior is an advantage when modeling amplifiers operating with non-Gaussian input signals. The recurrence relation for Chebyshev polynomials is as (4.9):

$$T_{n+1}[u(k)] = 2u(k)T_n[u(k)] - T_{n-1}[u(k)] \tag{4.9}$$

The first few Chebyshev polynomials of the first kind are:

$$T_0[u(k)] = 1 \tag{4.10}$$
$$T_1[u(k)] = u(k) \tag{4.11}$$
$$T_2[u(k)] = 2u^2(k) - 1 \tag{4.12}$$
$$T_3[u(k)] = 4u^3(k) - 3u(k) \tag{4.13}$$

Hermite Polynomials

If the input signal is normalized as a zero-mean, unit-variance Gaussian distribution, the Hermite polynomials are the choice for the regression matrix orthogonalization.

The recurrence formula for Hermite polynomials is [51]:

$$H_{n+1}[u(k)] = 2u(k)H_n[u(k)] - 2u(k)H_{n-1}[u(k)] \tag{4.14}$$

The first Hermite polynomials are:

$$\begin{aligned} H_0[u(k)] &= 1 & (4.15)\\ H_1[u(k)] &= u(k) & (4.16)\\ H_2[u(k)] &= u^2(k) - 1 & (4.17)\\ H_3[u(k)] &= u^3(k) - 3u(k) & (4.18) \end{aligned}$$

4.2.2 Baseband Power Series

Although polynomial LS estimation is a reasonable possibility to calculate the IMD components, it generates also "out-of-band" harmonics (second, third harmonic zone and so on), as shown as example for a two-tone excitation in Fig. 4.3.

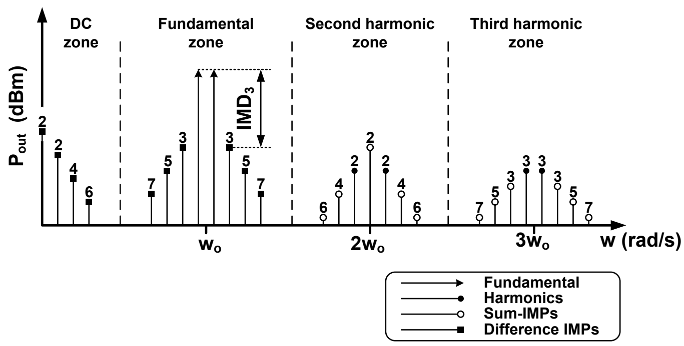

Figure 4.3: *Frequency-domain response of a nonlinear amplifier supplied with a two-tone test input signal.*

These are uninteresting for pre-distortion purposes, the main objective of behavioral modeling. To solve this problem, the first-zone equivalent (or baseband) polynomial is necessary. It can be derived writing the input signal as [55]:

$$\begin{aligned} u(k) &= Re\left[u(k)e^{jwk}\right] & (4.19)\\ &= \frac{1}{2}\left[u(k)e^{jwk} + u(k)^* e^{-jwk}\right] & (4.20) \end{aligned}$$

So, a binomial based expression for $u^n(k)$ can be obtained:

$$\begin{aligned} u^p(k) &= \left\{\frac{1}{2}\left[u(k)e^{jwk} + u(k)^* e^{-jwk}\right]\right\}^p & (4.21)\\ &= \frac{1}{2^p}\sum_{s=0}^{p}\binom{p}{s} u^p(k)u^{p-s}(k)e^{jw(2s-p)k} & (4.22) \end{aligned}$$

Only the terms where p is odd and $2s - p = \pm 1$ can contribute for the first-zone output, or $s = (p+1)/2$ and $s = (p-1)/2$. Then (4.22) can be written as:

$$\begin{aligned} u_{1z}^p = &\frac{1}{2^p}\sum_{s=0}^{p}\binom{p}{\frac{p+1}{2}}[u(k)]^{\frac{p+1}{2}}[u^*(k)]^{\frac{p-1}{2}}e^{jwk}+ \\ &+\frac{1}{2^p}\sum_{s=0}^{p}\binom{p}{\frac{p-1}{2}}[u(k)]^{\frac{p-1}{2}}[u^*(k)]^{\frac{p+1}{2}}e^{-jwk} \end{aligned} \tag{4.23}$$

Using the binomial property and the relation observed in (4.20)

$$u_{1z}^p = \frac{1}{2^{p-1}}\binom{p}{\frac{p+1}{2}}[u(t)]^{\frac{p+1}{2}}[u^*(k)]^{\frac{p-1}{2}} \tag{4.24}$$

Finally the first-zone filtered input signal can be found as:

$$u_{1z}^p = \frac{1}{2^{p-1}}\binom{p}{\frac{p+1}{2}}u(k)|u(k)|^{p-1} \tag{4.25}$$

The component $\frac{1}{2^{p-1}}\binom{p}{\frac{p+1}{2}}$ corresponds to the baseband power series coefficients, and can be determined as proposed in the next section.

Baseband Polynomial LS Estimation

If no bias is present in the input/output signals, the regression matrix can be written as:

$$\mathbf{U} = \begin{bmatrix} u(1) & u(1)|u(1)|^2 & \dots & u(1)|u(1)|^{p-1} \\ u(2) & u(2)|u(2)|^2 & \dots & u(2)|u(2)|^{p-1} \\ \vdots & \vdots & \ddots & \vdots \\ u(N) & u(N)|u(N)|^2 & \dots & u(N)|u(N)|^{p-1} \end{bmatrix} \tag{4.26}$$

The baseband polynomial can be written in a compact form:

$$\phi_p[u(k)] = \sum_{p=1}^{P} c_p|u(k)|^{2(p-1)}u(k) \tag{4.27}$$

Pre-Processing for Baseband Polynomials

An alternative way to determine baseband polynomials is using a pre-processing method. This procedure divides the measured data in AM/AM and AM/PM conversion curves in slices dependent of the input power and the number of points involved. Figure 4.4 shows a histogram that contains the number of points taken into account to construct the corresponding slice. This histogram shows the irregular distribution of the points, dependent on the instantaneous input power. Then the signal AM/AM and AM/PM conversion curves are determined directly from the mean of the slices, as in Fig. 4.5. From these curves, a baseband polynomial can be fitted and the amplifier nonlinearity is then parameterized.

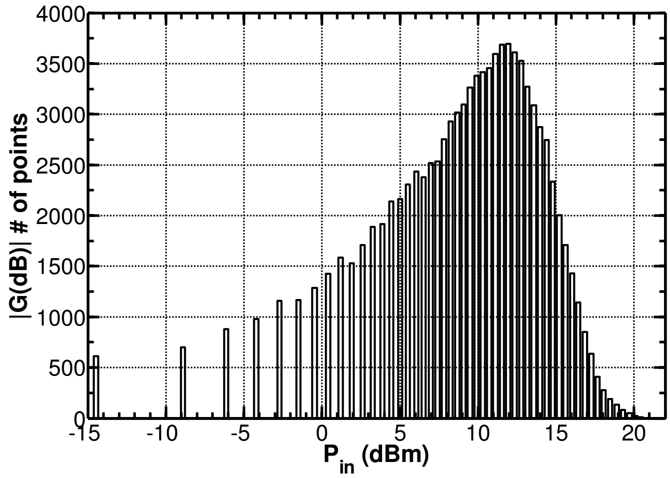

Figure 4.4: *Histogram of the AM/AM nonparametric estimation.*

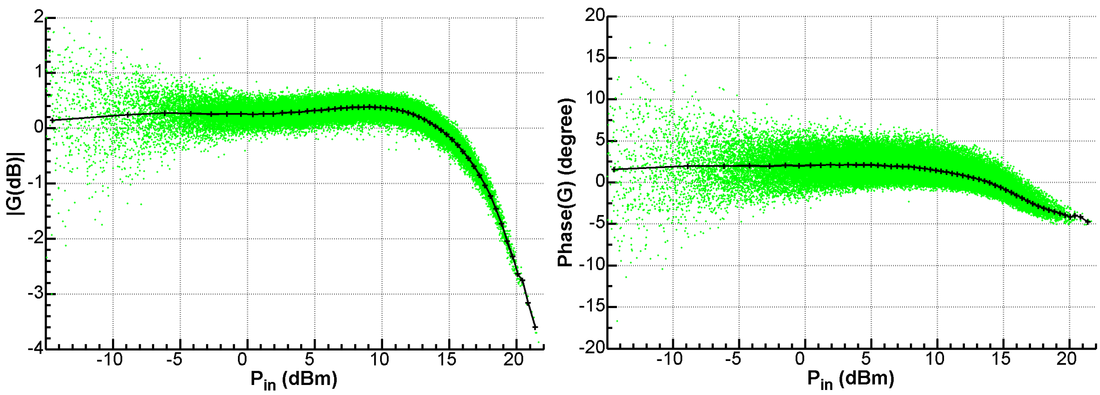

Figure 4.5: *AM/AM and AM/PM pre-processing.*

Raich-Zhou Polynomials

These polynomials are orthogonal for a baseband complex Gaussian input, presenting a numerical improvement in the Hessian matrix CN in comparison with the power series polynomial. It has a closed-form expression and is especially designed for baseband BM estimation, using unity variance complex input signals. All these properties are an advantage in comparison with other orthogonal approaches like G-functionals [18] or Hermite polynomial. A detailed proof and derivation of these polynomials can be found in [28]. The polynomials until the 9^{th} order for the conventional polynomial basis function as in (4.27) are listed below:

$$\phi_1 = u(k) \tag{4.28}$$

$$\phi_3 = \sqrt{2}\left(-1 + \frac{1}{2}|u(k)|^2\right)u(k) \tag{4.29}$$

$$\phi_5 = \sqrt{3}\left(1 - |u(k)|^2 + \frac{1}{6}|u(k)|^4\right)u(k) \tag{4.30}$$

$$\phi_7 = \sqrt{4}\left(-1 + \frac{3}{2}|u(k)|^2 - \frac{1}{2}|u(k)|^4 + \frac{1}{24}|u(k)|^6\right)u(k) \tag{4.31}$$

$$\phi_9 = \sqrt{5}\left(1 - 2|u(k)|^2 + |u(k)|^4 - \frac{1}{6}|u(k)|^6 + \frac{1}{120}|u(k)|^8\right)u(k) \tag{4.32}$$

With a recurrence equation:

$$\Psi_{2w+1}(u(k)) = \sum_{s=0}^{w}(-1)^{w-s}\frac{\sqrt{w+1}}{(s+1)!}\binom{w}{s}\phi_{2s+1}(u(k)) \tag{4.33}$$

The same authors also present an orthogonal polynomial for uniformly distributed input signals in the interval $[0,1]$, with odd and even terms (see [29]).

4.2.3 Saleh Model

The Saleh memoryless nonlinear model was widely applied for Traveling Wave Tube Amplifier (TWTA) models. It is characterized by two equations, one for the AM/AM characteristics and another one for the AM/PM, as following

$$f_{\mathrm{AM/AM}} = \frac{\alpha_1|u(k)|}{1+\beta_1 u(k)^2} \tag{4.34}$$

$$f_{\mathrm{AM/PM}} = \frac{\alpha_2|u(k)|^2}{1+\beta_2|u(k)|^2} \tag{4.35}$$

The parameters α_1, α_2, β_1 and β_2 completely characterize the model. Its fundamentals and estimation techniques can be found in [58]. Figure 4.6 presents a block diagram of this model.

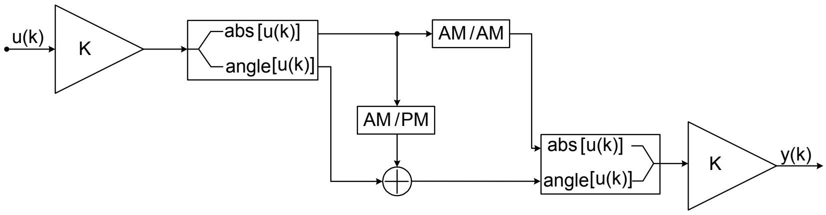

Figure 4.6: *Saleh model block diagram.*

4.2.4 Neural Networks (NN)

Another possible method to estimate an amplifier nonlinearity is to use neural networks (NN). Although the estimation with NN is an interesting alternative, care should be taken not to switch from a completely unknown model (amplifier) to another one (very complex NN structure), due to the low flexibility and information that can be extracted from a very complex NN structure.

Neural Networks Estimation

The complex valued signal is split in real and imaginary parts, for the use in Matlab® NN algorithm [49]. So the input and output layers have 2 inputs and consequently 2 neurons.

The network topology was composed of linear activation functions in the input (compare (4.36)), followed by tangent-sigmoid activation functions in the hidden layer (compare (4.37)), and linear activation functions on the output.

$$\mathrm{lin}[u(k)] = u(k) \tag{4.36}$$

$$\mathrm{tansig}[u(k)] = \frac{2}{1 + e^{-2u(k)} - 1} \tag{4.37}$$

The network configuration can be seen in Fig. 4.7.

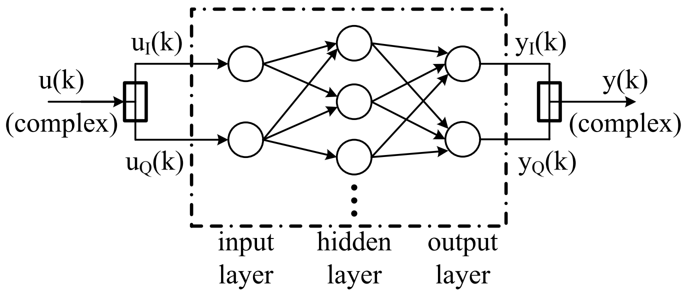

Figure 4.7: *NN configuration for a memoryless estimation.*

The main drawbacks of NN are:

- The number of neurons used should be previously known, or the network has to be pruned or augmented and recalculated;

- The weights are internal to the network and do not always have a direct correspondence to Volterra kernels as polynomials;
- The tangent-sigmoid activation function generates only odd-order harmonics, but not only on baseband (first-zone filtered), as baseband polynomials.
- The network is nonlinear in parameters, so the method used for parametrization should be also capable of nonlinear estimation. The complexity is clearly increased;
- With a very high complexity, NN have low possibility to be implemented in DSP's, or embedded systems with reduced computational capacity.

The advantages are:

- Having a higher complexity and using nonlinear methods for the system estimation, NN can sometimes achieve better results than the estimation using polynomials;
- The function used (tangent-sigmoid) is bounded, eliminating any divergence possibilities.

4.2.5 Look-Up Tables

Look-up tables (LUTs) are the most common type of nonlinear static models in real-world implementations [40]. An advantage in comparison with other methods is the configuration possibility of the interpolation and extrapolation behavior. LUTs also present good accuracy, and very fast evaluation. The drawbacks are: poor physical interpretation, high number of parameters, and not continuously differentiable. For the special case of PAs, LUTs can be parametrized using the technique already explained in section 4.2.2, from AM/AM and AM/PM characteristic curves. Linear interpolation is normally used to determine the points among intervals, but also other methods as cubic interpolation and splines are possible [49].

4.3 Linear Memory

The two-box modeling technique is a possibility to represent the linear memory of an amplifier.

4.3.1 Two-Box Models

Two-box models are also known as modular approach [24] or feed-forward block oriented models [25]. They are obtained by combining components from the following two classes: static (or memoryless) nonlinearities and causal, linear time-invariant dynamic subsystems. Parametric and nonparametric modeling methodologies can be used. Flexible arrangements of block structured models in two possibilities are feasible: Wiener model (Linear-Nonlinear) and Hammerstein model (Nonlinear-Linear) [20].

The most frequently used configuration is a FIR filter and a nonlinearity, represented by a polynomial [25]. Examples of these structures are shown in Fig. 4.8.

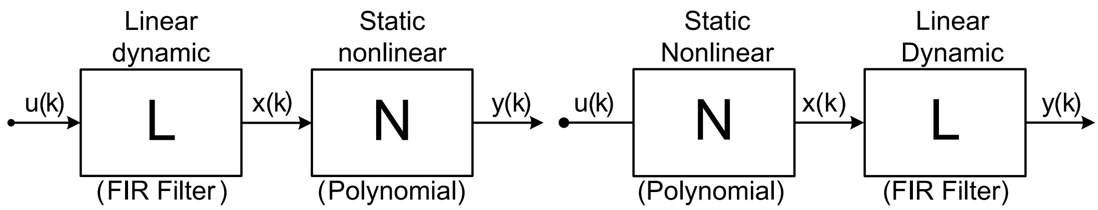

Figure 4.8: *Wiener (left) and Hammerstein (right) models.*

If the linear dynamic block is represented by a FIR filter, the output of this block for the Wiener Model is:

$$y_{\mathrm{WL}}(k) = \sum_{\tau=0}^{M-1} h(\tau)u(k-\tau) \tag{4.38}$$

For the Hammerstein Model, the FIR filter output is:

$$y_{\mathrm{HL}}(k) = \sum_{\tau=0}^{M-1} h(\tau)x(k-\tau) \tag{4.39}$$

If the static nonlinearity block is represented by a power series, the output of this block can be formulated for the Wiener Model as (4.40) and as (4.41) for the Hammerstein Model.

$$y_{\mathrm{WNL}}(k) = \sum_{p=0}^{P} c_p x^p(k) \tag{4.40}$$

$$y_{\mathrm{HNL}}(k) = \sum_{p=0}^{P} c_p u^p(k) \tag{4.41}$$

The overall model output is then the combination of these equations for each model:

$$y_{\mathrm{W}}(k) = \sum_{p=0}^{P} c_p \left(\sum_{\tau=0}^{M-1} h(\tau)u(k-\tau) \right)^p \tag{4.42}$$

$$y_{\mathrm{H}}(k) = \sum_{\tau=0}^{M-1} h(\tau) \left(\sum_{p=0}^{P} c_p u^p(k-\tau) \right) \tag{4.43}$$

Equations (4.42) and (4.43) are a simple way to model a nonlinear amplifier with memory.

4.3.2 Volterra Kernels Relationship to Wiener and Hammerstein Models

As Volterra kernels are the complete and reliable descriptors of the system's behavior response [24], it is interesting to find the relationship between Wiener and Hammerstein models and Volterra kernels. The finite VS model is given by (4.1).

The p^{th} order kernel describes nonlinear interactions among p copies of the input. For $p = 0$ a constant output is the response, for $p = 1$ the one dimension linear kernel will be the output, for $p = 2$ the nonlinear interactions between two copies of the input will result in two-dimension matrix and so on.

If equation (4.42) is written as:

$$y_{\mathrm{W}}(k) = \sum_{p=1}^{P} c_p \left(\sum_{\tau_1=0}^{M-1} \cdots \sum_{\tau_q=0}^{M-1} h(\tau_1) \ldots h(\tau_q) u(k-\tau_1) \ldots u(t-\tau_q) \right) \tag{4.44}$$

The resulting relationship for the order $p^t h$ order Volterra kernel will be as follows [25]:

$$h_p(\tau_1, \cdots, \tau_q) = c_p h(\tau_1) h(\tau_2) \ldots h(\tau_q) \tag{4.45}$$

So the order p Volterra kernel is equal to the product of p copies of the impulse response function of the linear block multiplied by the p^{th} nonlinearity coefficient.

To verify if a system can be represented by a Wiener model it is necessary that any one-dimensional slice, taken parallel to an axis of a Volterra kernel would be proportional to the first order kernel, as stated mathematically in (4.45). This is not a sufficient condition. This procedure is described in [59]. This condition means also that for the Wiener model estimation, all input signal combinations cannot be taken into consideration.

An example of a 2^{nd} order Volterra kernel of a Wiener system is shown in Fig. 4.9.

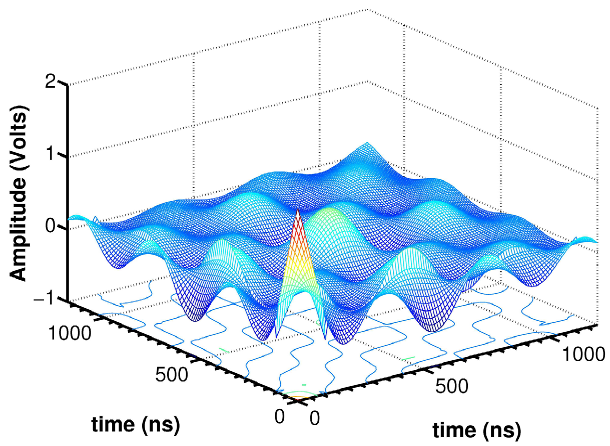

Figure 4.9: *Example of a hypothetical* 2^{nd} *order Volterra kernel of a Wiener system.*

The Hammerstein-Volterra relationship can be derived in a similar way. The equivalent Volterra kernels are as in (4.46).

$$h_p(\tau_1, \cdots, \tau_q) = \begin{cases} c_p h(\tau) & \tau_1 = \tau_2 = \cdots = \tau_q \\ 0 & \text{otherwise} \end{cases} \tag{4.46}$$

The Volterra kernels of a Hammerstein system are only nonzero along their diagonals ($\tau_1 = \tau_2 = \cdots = \tau_q$).

A system can be treated as a Hammerstein model if the kernels are only nonzero along their diagonals, and is also necessary that these diagonals are proportional to the first order kernel [59].

An example of a 2^{nd} order Volterra kernel of a Hammerstein system is shown in Fig. 4.10.

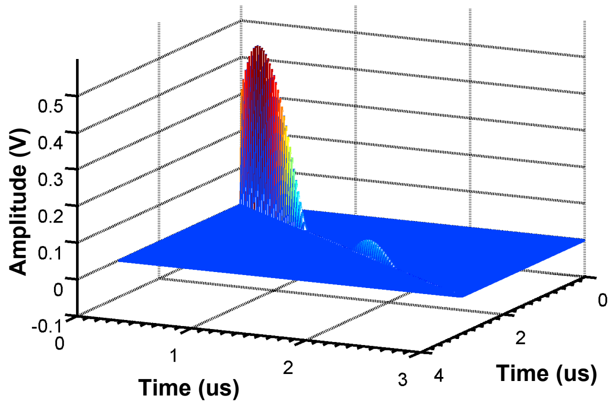

Figure 4.10: *Example of a hypothetical 2^{nd} order Volterra kernel of a Hammerstein system.*

Equations (4.45) and (4.46) show how close these representations are to VS, with the advantage that Wiener and Hammerstein models do not need large matrices inversions required by Volterra models in the estimation process. Fifth order models or even seventh order models can be represented in a very parsimony way.

4.3.3 Wiener and Hammerstein Models Estimation

A procedure to estimate a Wiener model is:

1. Parameterize a FIR filter estimated from the measured input/output signal, using LS combined with SVD techniques [13];

2. Fit a nonlinearity between the filtered input signal ($x(k)$ in Fig. 4.8) and the output signal, using LS for polynomials.

An alternative way is:

1. Fit a nonlinearity between the measured input/output signal, using LS for polynomials;

2. Estimate the intermediate signal ($x(k)$ in Fig. 4.8): apply LS for polynomials between the measured output/input signal. In this way, it is possible to fit the polynomial inverse. Then $x(k)$ is calculated applying the measured output signal to this inverse polynomial;

3. Identify a FIR filter between the measured input signal and the intermediate signal $x(k)$.

An algorithm to estimate a Hammerstein model is:

1. Fit a nonlinearity between the measured input/output signal, using LS for polynomials;

2. Estimate a FIR filter between the measured input signal and the intermediate signal $x(k)$.

Estimating Wiener and Hammerstein models using this two-step LS procedure is not the most accurate method, because the influence of a component of the model (filter or polynomial) is always neglected for the first estimation. Nevertheless, the method is linear in parameters, allows a first estimation of the system memory, and uses less parameters than other methods (described in next section).

4.4 Nonlinear Memory

More complex models are necessary to estimate the nonlinear memory, like parallel models or a close approximation of the VS model, the Wiener-Bose model.

4.4.1 Parallel Cascade Models

Any system that can be represented by a truncated VS (4.1) can be also modeled exactly using parallel cascaded structures [60].

This technique is the association in branches of various models (Wiener, Hammerstein, Wiener-Hammerstein etc.). The overall model structure becomes more complicated with each iteration, as each branch is composed by a single model. The value of the cost function decreases or stays constant with each additional branch [61]. An example of this configuration is seen in Fig. 4.11. This method combines following fa-

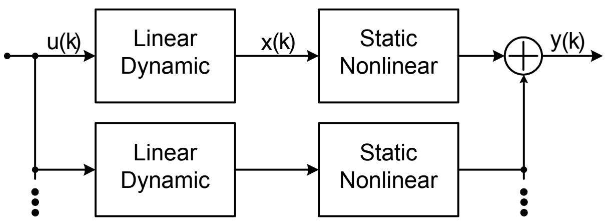

Figure 4.11: *Example of a parallel Wiener model.*

vorable properties: It is computationally efficient even for high-order models with large memory-bandwidth products, allows the direct extraction of the Volterra kernels, and offers the convenience to use different methods for the identification of the linear and nonlinear blocks [61]. But is very sensitive to noise if too many paths are used [24]. Consequently, a proper selection of the paths using parametric FOMs and the system order of the nonlinearity should be made to achieve low noise and good convergence models.

4.4.2 Parallel Cascade Models Estimation

The general algorithm is [30, 61], and the procedure is represented graphically for a parallel Wiener model as shown in Fig. 4.12:

1. Fit the first cascade from amplifier measurement input/output data;

2. Compute the output from the first branch and subtract it from the measured output to find the first residue: $\mathrm{res}_1(k) = y_{\mathrm{meas}}(k) - y_{\mathrm{m1}}(k)$;

3. Fit a second branch between the input and the first residue (Intermediate black-box model);

4. Compute the second residue subtracting the output of the second path from the first residue: $\mathrm{res}_2(k) = \mathrm{res}_1(k) - y_{\mathrm{m}2}(k)$;

5. Include this branch in the model only if FOM results are better than previous results;

6. Continue this procedure until no improvements in the chosen FOM are obtained.

The model has its output defined as:

$$y_{\mathrm{PW}}(k) = \sum_{i=1}^{I} \sum_{p=0}^{P} c_p^{(i)} \left(\sum_{\tau=0}^{M-1} h^{(i)}(\tau) u(k-\tau) \right)^p \tag{4.47}$$

where I is the total number of paths, P denotes the maximum order of the polynomial used, $c_p^{(i)}$ are the polynomial coefficients for the i^{th} path, M is the memory length, k and τ are discrete indexes of the sampling interval, and $h^{(i)}(\tau)$ is the impulse response i^{th} path.

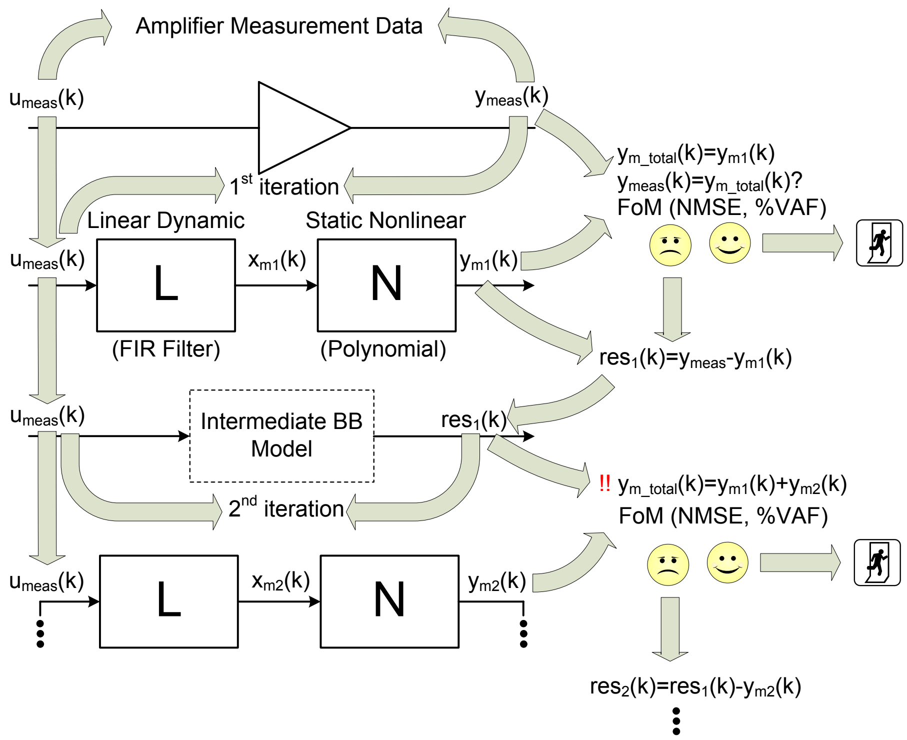

Figure 4.12: *Graphical parallel Wiener algorithm.*

4.4.3 Wiener-Bose Model

Models that can approximate the behavior of VS were studied and implemented in different areas as physiology [23] and automotive [40]. The Wiener-Bose architecture

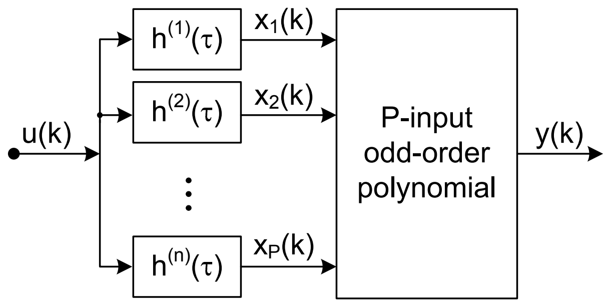

Figure 4.13: *Wiener-Bose model structure.*

(compare structure in Fig. 4.13) is valid for systems with fading memory (i. e., systems which Volterra kernels are absolutely summable on the system memory $[0, M]$), as a PA [62].

It combines a delay structure with a multi-input polynomial, providing every possible combination of the input signal. This model was also called VS approximation [62, 63], and is a powerful tool to estimate nonlinear systems.

The output of the i^{th} filter in the bank is [23]:

$$x_i(k) = \sum_{\tau=0}^{M-1} h^{(i)}(\tau) u(k-\tau) \tag{4.48}$$

The static nonlinearity is formulated as a multiple input polynomial of degree P:

$$m(x_1, \ldots, x_P) = \sum_{p=0}^{P} \sum_{i_1=1}^{Q} \sum_{i_2=i_1}^{Q} \cdots \sum_{i_q=i_{q-1}}^{Q} c_p^{(i_1,i_2,\ldots,i_p)} M_p(x_{i_1}(\tau), x_{i_2}(\tau), \ldots, x_{i_p}(\tau)) \tag{4.49}$$

where $M_p(x_{i_1}(\tau), x_{i_2}(\tau), \ldots, x_{i_p}(\tau))$ is the product of the arguments x_1 through x_P of the p^{th} order multiple input polynomial.

Unifying these equations, the overall output of the Wiener-Bose model is:

$$y(k) = \sum_{p=0}^{P} \sum_{i_1=1}^{Q} \sum_{i_2=i_1}^{Q} \cdots \sum_{i_p=i_{p-1}}^{Q} c_p^{(i_1,i_2,\ldots,i_p)} x_{i_1}(\tau) x_{i_2}(\tau) \ldots x_{i_p}(\tau) \tag{4.50}$$

The Wiener-Bose model has a general output equation and can be estimated in many different ways. One of these possibilities is described in the next chapter.

4.5 Conclusion

A power amplifier is a mild nonlinear system, that can be fully described with a VS, although building the former requires a very high number of coefficients. The linear estimation technique has the most suitable characteristics for amplifier BMs estimation. Depending on the memory presented by the amplifier, different amplifier models can be used: memoryless, linear memory and nonlinear memory. The parallel models and the Wiener-Bose Volterra approximation are powerful methods to estimate an amplifier behavioral model. For pre-distortion purposes, the models presented can be inverted using the output signal as input signal, and input signal as output signal in the LS estimation.

Chapter 5

Practical Model Identification

5.1 Introduction

This chapter describes measurement signals and setups, different estimation techniques, and selected applications of BMs. It is dedicated to the search of efficient and accurate BMs for pre-distortion purposes.

5.2 Measurements

Behavioral models are fully based on measurements. This section will present PAs and input signals characteristics used to generate the input/output measurement signals, referenced in this entire chapter.

5.2.1 Measured PA's

Two amplifiers were used for measurements. The first one is an Ericsson base station PA (amplifier A), and the second one is a Freescale Board, designed for Wimax applications (amplifier B).

Amplifier A

The class AB main amplifier had the following nominal characteristics: Frequency range of 1.93 to 1.99 GHz, maximum output power of +48 dBm, 36 dB gain, and 1 dB output compression point of +53 dBm. This amplifier uses a Motorola 90 W MRF 18090A LDMOS transistor in the final stage. The amplifier output is protected by an isolator which assures proper matching and negligible reverse gain.

An important issue for the amplifier measurements was to assure thermal stability. The amplifier was cooled by a fan producing a flow rate of $95\,\mathrm{m}^3/\mathrm{h}$ which set the amplifier casing temperature at approximately 40 °C at maximum output power. The measurements were performed starting from the linear and moving to the compression operation region. Before taking a measurement at a fixed input back-off (IBO) the DUT was driven for one hour by the input signal.

The measured amplifier gain curves are shown in Fig. 5.1. The gain magnitude and phase present significant changes in the area near the 1 dB compression point.

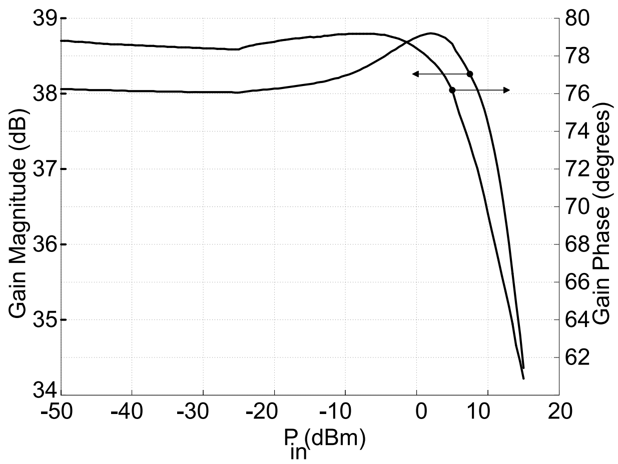

Figure 5.1: *Amplifier A gain curves.*

Amplifier B

The amplifier was a Freescale board equipped with a MRFS3801H LDMOS transistor. This 10 W transistor is optimized for base station applications up to 3.8 GHz. It is suitable for WiMax, WiBro and OFDM multicarrier Class AB and Class C amplifier applications. The drain bias circuit of amplifier board was modified to introduce the memory effects, as the amplifier board from Freescale shows a static but not a dynamic nonlinear behavior. The gain curve is displayed in Fig. 5.2.

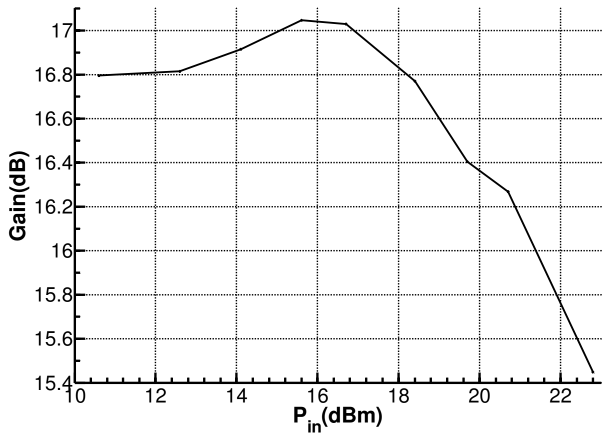

Figure 5.2: *Amplifier B gain curve.*

5.2.2 Excitation Signals and Measurement Setups

Different measurement setups were used for the various excitation signals.

QAM Signal

A potential and attractive modulation scheme when high data rate transmission is needed is the 16-QAM. It offers high power/spectral efficiency [64], but lacks a constant envelope and becomes highly sensitive to PA nonlinearities.

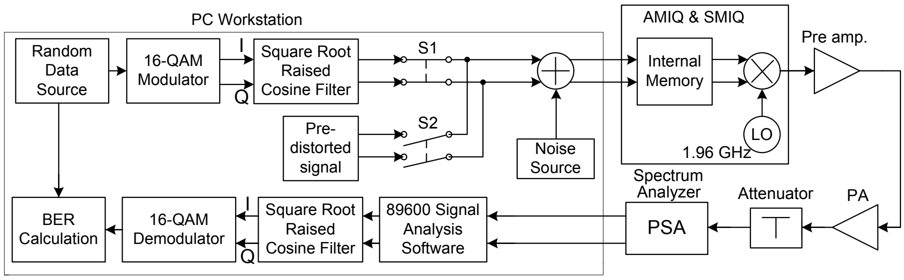

Figure 5.3: *Measurement setup for modeling (S1 closed and S2 open) and pre-distortion procedures (S1 open and S2 closed).*

The measurement system is presented in Fig. 5.3. The 16-QAM modulated input signal was filtered by a root raised cosine (RRC) filter with roll-off equal to 0.25, loaded in the I/Q Modulation Generator Rohde & Schwarz AMIQ, and upconverted by a Vector Signal Generator Rohde & Schwarz SMIQ. By varying the noise source level (compare Fig. 5.3), it was possible to generate measurement signals with different levels of SNR. The RF bandwidth of the signal was 2.5 MHz. The PA's output signal was measured at a center frequency of 1.96 GHz using an Agilent Performance Spectrum Analyzer (PSA) operating at a sampling rate of 10 MSamples/s, and was processed by the Agilent 89600 Signal Analysis Software.

Noise Source

The noise source was implemented to corrupt the amplifier input signal (and consequently the output signal) with Gaussian noise.

In order to quantify the noise level, the Signal to Noise Ratio (SNR in dB) was used as a parameter. It is defined as the ratio of the signal power S to the noise power N and is related to $\frac{E_b}{N_0}$ in dB for complex signals as following [49]:

$$\mathrm{SNR} = 10\ log\left(\frac{E_\mathrm{b}}{N_0}\right) + 10\ log(k) - 10\ log\left(\frac{T_\mathrm{sym}}{T_\mathrm{samp}}\right) \tag{5.1}$$

where $\frac{E_\mathrm{b}}{N_0}$ is the ratio of bit energy to noise power spectral density, k is the number of information bits per symbol, T_sym is the signal symbol period and T_samp is the signal sampling period. The noise bandwidth is equal to the sampling frequency.

WCDMA Signal

To estimate BMs for WCDMA amplifiers, a WCDMA signal with 3.84 MHz RF bandwidth, 10 dB Peak-to-Average Power Ratio (PAPR), and 12.2 kbps data rate was used. The PA was at operating at 5 dB IBO. The Vector Signal Analyzer was operating at a sampling rate of 35 MSamples/s with a SNR of 60 dB for this input signal. In Fig. 5.4 the power spectral density of the measured input/output signal is shown at compression. Figure 5.5 shows the block diagram of the measurement setup.

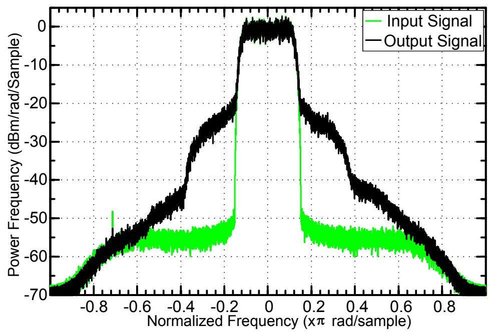

Figure 5.4: *PSD of the WCDMA input/output measured signal.*

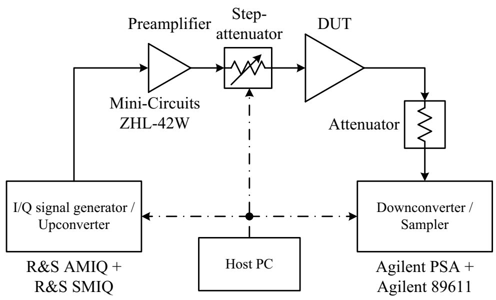

Figure 5.5: *Measurement system used to obtain the RF PA input/output data.*

WCDMA Signal – Simulation

In order to obtain perfectly synchronized and noiseless signals for BMs estimation, a simulation was performed. The sets of data are from an Advanced Design Systems PA model (compare Fig. 5.6), where different memory effects (memoryless (noM), linear memory (LM), nonlinear memory (NLM) and linear and nonlinear memory (LM/NLM)) could be selected changing the characteristics of the sub-circuits SC1 and SC2. Four sets of WCDMA simulated data (same characteristics as described above), acquired at 2 dB input back-off (IBO), were used for model extraction and validation. Figure 5.7 shows AM/AM characteristics for these signals.

Wimax Signal

To estimate BMs using Wimax signal, the Wimax excitation signal was generated by the use of the Visual System Simulator from AWR, according to the following specification (based on [65]): WirelessMAN-OFDM (256-carriers), bandwidth of 3.5 MHz, 64 QAM, code rate of 3/4, length of cycling prefix of 1/8, and sampling factor of 8/7. The final input signal resulted in a Peak-to-Average Power Ratio (PAPR) of 9.7 dB.

The signal was captured at a rate of 28 MSamples/s, which corresponds to a measurement duration of 2.5 ms, using the measurement setup from Fig. 5.5 at 3.5 GHz. The amplifier output power was 35.6 dBm (2 dB IBO).

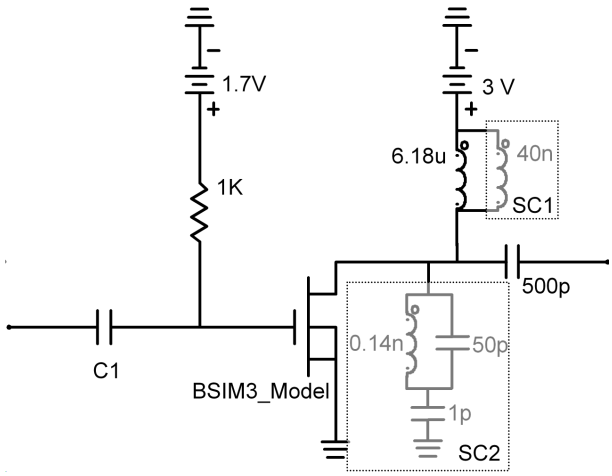

Figure 5.6: *ADS equivalent circuit model used in simulations.*

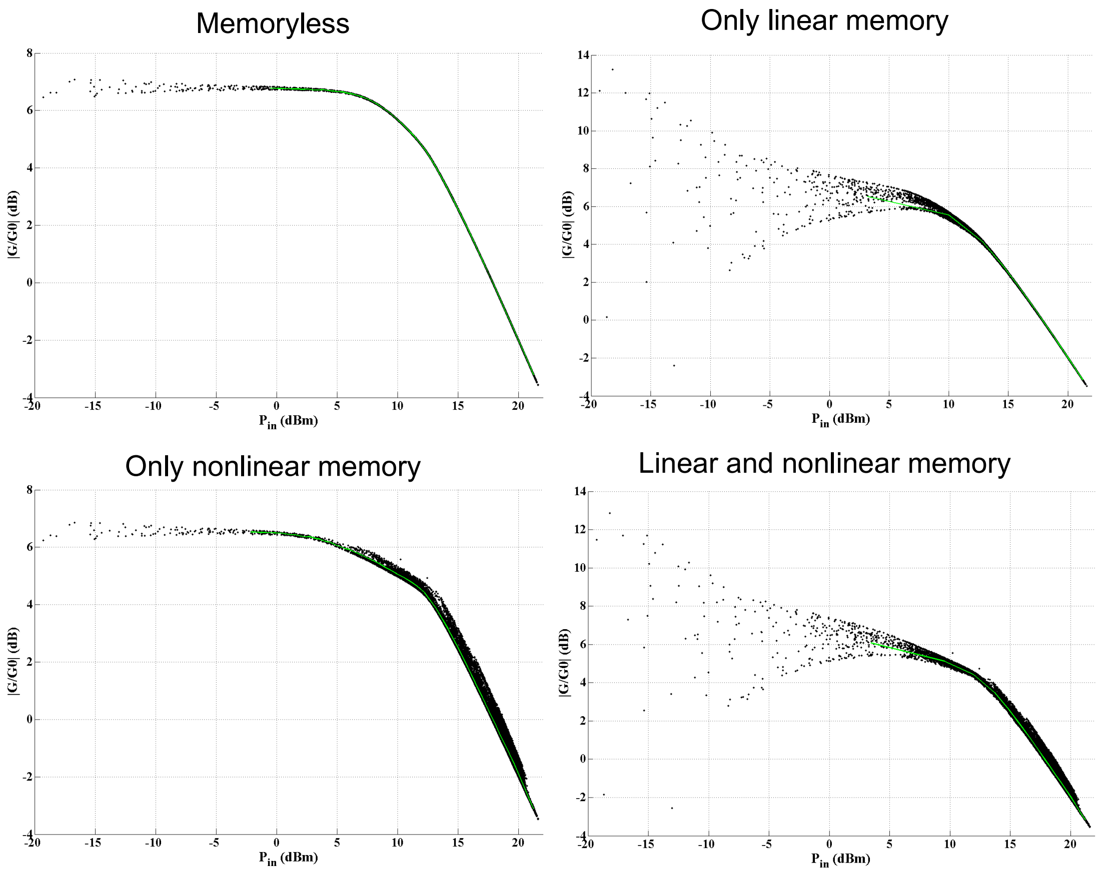

Figure 5.7: *AM/AM characteristics of the signals presenting different types of memories.*

5.3 Identification in the Presence of Noise

This section discusses the black-box modeling of a PA using a quadrature amplitude modulated (QAM) input signal corrupted by different levels of Additive White Gaussian Noise (AWGN).

5.3.1 Introduction

Understanding the influence of AWGN in the PA modeling can improve the identification methods, as noise is always present in the measured input/output signals. In [66] an interesting investigation of several 16-QAM constellations in a noisy environment over a satellite channel in the linear and nonlinear case was presented. In [67] three PA models were simulated and analyzed with analytical expressions using CDMA modulations with AWGN and different back-off levels, although no amplifier modeling technique was involved. Kumar [68] showed simulations of a PA with 16-QAM signals and some back-off configurations, but the amplifier is approximated as a limiter. A study of AWGN influence in the estimated models can also help to decide whether to use noise free data (NFD) models or noise corrupted data (NCD) models. The denomination NFD is relative to the digital generation of the signal, and not to the measured signal, that indeed contains measurement noise.

This kind of identification has two main difficulties:

- The nonlinear output signal components generated by the amplifier at low input back-off (IBO) operating conditions can change the output constellation and lead to difficulties in finding the correct model parameters;

- The AWGN degrades the SNR and can lower the model accuracy that should ideally not be influenced.

The IBO is here specified as the average input power level with respect to the 1 dB compression point at the input.

Firstly, the models were estimated using NFD and then with different levels of NCD. NFD was also applied to the models calculated with NCD and vice versa. Two situations were considered: 6 dB IBO and 2 dB IBO.

Final results show a loss of accuracy caused by noise in the identified models. The achieved modeling results were analyzed to check how an estimated model is affected by noise added to the measured input/output signals.

5.3.2 Measurement Setup

The PA used (amplifier A) is described in section 5.2.1. The excitation signal is described section 5.2.2 (QAM signal), as well as the setup used to obtain the measured input/output signals with different levels of SNR.

5.3.3 Amplifier Identification and Investigation

By using the SVD technique (compare 3.2.3) Wiener models were identified from the measurement results. All models were calculated using a small segment (30%) of the recorded input/output signals at different IBO levels and were cross-validated with the remaining data set until a satisfactory %VAF (or NMSE) was achieved. The estimated FIR filter had 3 taps and the model nonlinearity was implemented by a 9^{th} order

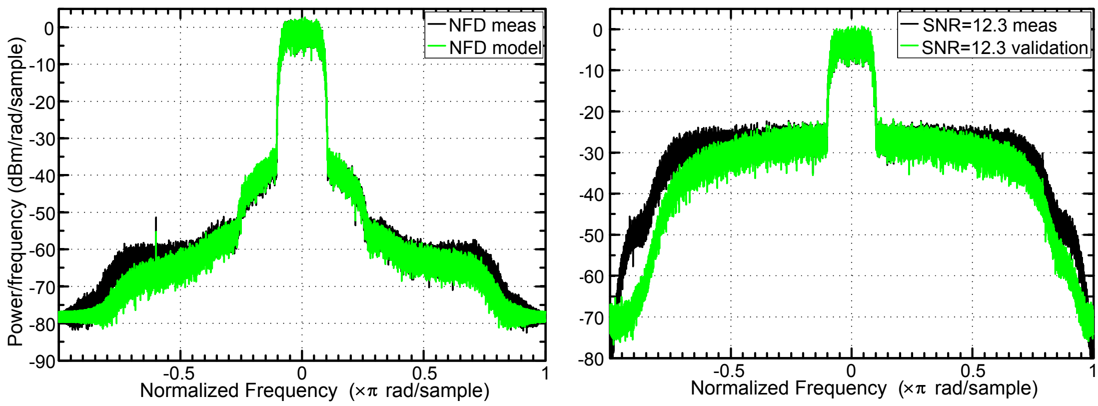

Figure 5.8: *PSD curves for the NFD estimated model at* 6 dB *IBO (left), and PSD curves for the validation data SNR* $= 12.3$ dB *applied to the same model (right).*

polynomial. The process used in BM analysis is demonstrated in Table 5.1. The NFD models were tested with NFD (case A) and NCD (case B); NCD models were tested with NFD (case C) and NCD (case D).

Table 5.1: *Process used for BM analysis with different signals*

		Validation	
	Models	NFD	NCD
Modeling	NFD	A	B
	NCD	C	D

5.3.4 Noise Free Amplifier Identification

The models were first estimated using NFD. In this ideal situation the noise source was inactive and identification at different IBO levels was realized. These results are shown in the first column of Tables 5.2 and 5.3, under the NFD section. Then NFD models were tested with the same signals contaminated with noise (NCD case). Two levels of SNR were used. The results are listed in the second and third column and show that these models have a high sensitivity to noise. PSD curves from a NFD model estimated at 6 dB IBO and validated with NFD, and PSD curves for validation data SNR $= 12.3$ dB applied to this NFD estimated model are depicted in Fig. 5.8.

Although the highest modeling accuracy is obtained when NFD are used as input/output signals, the models could not predict the noise influence, and in some cases the accuracy loss was as high as 18 dB (6 dB IBO model).

5.3.5 Noisy Data Amplifier Identification

In a second step, models were estimated using NCD with different levels of SNR. NFD was also applied to the models calculated with NCD. The results are represented in Table 5.2 and 5.3, section SNR $= 12.3$ dB and SNR $= 5.3$ dB. The column *model* shows the identification accuracy using only the NCD. The column *validation (NFD)* presents the results when NFD is applied to these models.

Table 5.2: *Summary of modeling and validation results,* 6 dB IBO

6 dB IBO							
	NFD			SNR = 12.3 dB		SNR = 5.3 dB	
	model	validation SNR 12.3 dB	validation SNR 5.3 dB	model	validation NFD	model	validation NFD
%VAF	99.9	97.5	97.6	99.8	99.2	99.7	99.2
NMSE	−34.0	−16.0	−16.0	−29.0	−21.0	−26.0	−21.0

Table 5.3: *Summary of modeling and validation results,* 2 dB IBO

2 dB IBO							
	NFD			SNR = 12.3 dB		SNR = 5.3 dB	
	model	validation SNR 12.3 dB	validation SNR 5.3 dB	model	validation NFD	model	validation NFD
%VAF	99.9	98.1	90.6	99.8	98.7	99.0	95.6
NMSE	−28.5	−17.0	−10.0	−26.0	−19.0	−20.0	−13.5

An example of this modeling procedure is shown in Figs. 5.9 and 5.10. These models were estimated at 6 dB IBO and 2 dB IBO, respectively, and validated using NFD.

A reasonable fit between measurement and model can be seen in Fig. 5.10, at the left side, but the same model fails to represent the nonlinearities 30 dB below the carrier signal for the NFD case, due to the high level of nonlinearities.

The models estimated with NCD showed a better performance when validated with NFD. Although estimated with NCD, they are more suitable to the general use, because of their reasonable results in comparison with NFD models and their ability to predict the noise that could be incident in the signals.

Some models identified with NCD in Table 5.2 and 5.3, section SNR = 12.3 dB for 2 dB IBO and 6 dB IBO had achieved better modeling accuracy than with NFD, and a

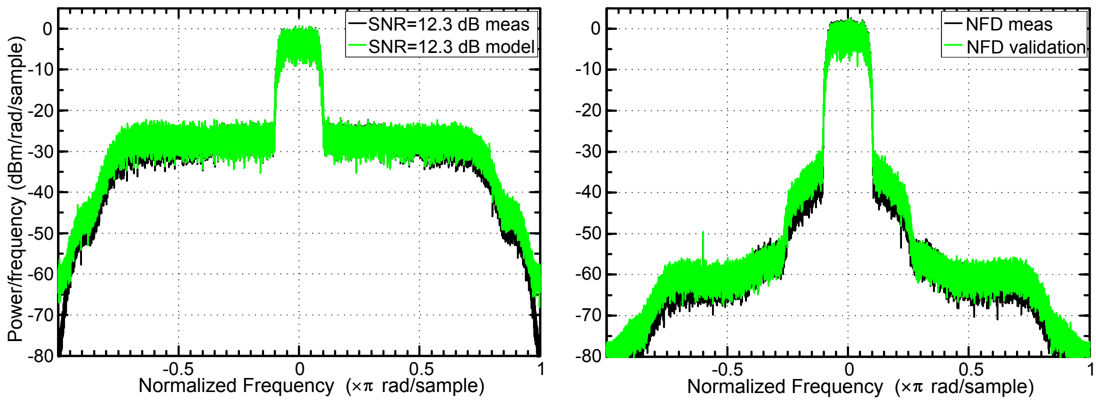

Figure 5.9: *PSD curves for the SNR* = 12.3 dB *estimated model at* 6 dB *IBO (left), and PSD curves for the NFD validation data applied to the same model (right).*

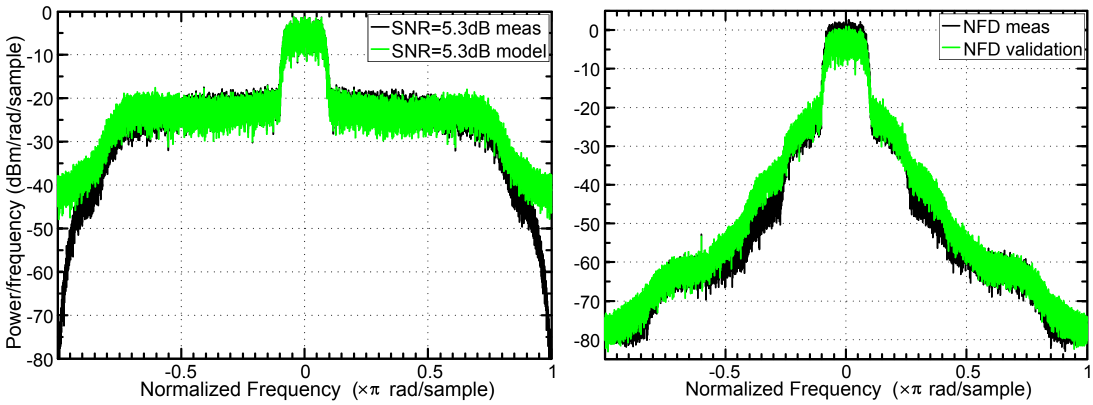

Figure 5.10: *PSD curves for the SNR* $= 5.3\,\mathrm{dB}$ *estimated model at* $2\,\mathrm{dB}$ *IBO (left), and PSD curves for the NFD validation data applied the same model (right).*

reasonable result when NFD was used. This can be explained by the fact that the most suitable signal to identify nonlinear models is the Gaussian noise [24], since it excites every possible source of nonlinearity, as already stated in chapter 2.

5.3.6 Conclusion

This section analyzed the influence of AWGN on a Wiener model using SVD techniques, with complex valued input/output signals. A comparison between models estimated with NFD and NCD was shown. The models estimated with NFD have an optimal performance only when NFD is used. The models estimated with NCD presented general superior results, once they account for the noise interference on the signals and have relatively close results to the NFD model. The SVD techniques exhibited reasonable results for PA modeling with 16-QAM modulation, even when AWGN was present.

5.4 A Hammerstein Pre-Distorter Based on a Wiener Model

This section will analyze the performance of a Hammerstein pre-distorter based on a Wiener model using SVD techniques, as this model showed good performance when AWGN is present.

5.4.1 Introduction

Among linearizing techniques proposed in the literature [1], digital pre-distortion (DPD) has become of significant importance and gained much attention because it offers reliability, flexibility and scalability in its implementation and performance. Among the different DPD techniques (RF, IF, Baseband), baseband DPD aims at compensating the whole transmitter in advance, and its methodology is independent of the type of PA used. DPD is currently carried out by means of high speed digital signal processors (DSP's, FPGA's), already present in most of the current communication equipment for mandatory issues within the wireless standards. In order to implement the DPD estimation process and later apply adaptive DPD to ensure real-time performance, a reliable PA nonlinear dynamic model is necessary.

A 16-QAM modulation scheme has been selected as input signal. Then SVD techniques are used to find a parsimonious Wiener model that is cross-validated with measurement data. Once the PA model is obtained, a Hammerstein based DPD is estimated by means of the indirect learning approach [69].

Simulation and experimental results are provided in order to show linearity improvement achieved by this Hammerstein DPD. In band and out-of-band distortion compensation was measured in terms of ACPR and EVM reduction.

5.4.2 Measurement Setup

The PA used (amplifier A) is described in section 5.2.1. The measured input/output signals were obtained from the measurement setup described in section 5.2.2 (QAM signal).

5.4.3 Digital Pre-Distorter Design

In order to pre-distort the PA characterized by a Wiener model, a Hammerstein model (depicted in Fig. 5.11) was considered. It is the structural inverse of a Wiener model. The memoryless nonlinearities have been modeled using baseband polynomial. For modeling the linear block, finite and infinite impulse response (FIR and IIR) filters have been applied. The Hammerstein model used for DPD is described in (5.2), and

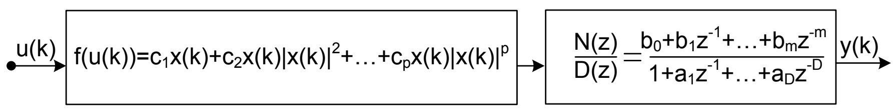

Figure 5.11: *Hammerstein nonlinear dynamic model.*

further details on the DPD *indirect learning* identification process can be found in chapter 1.

$$y(k) = \sum_{m=0}^{M-1} b_m \left(\sum_{p=0}^{P} c_p u(k-\tau_m)|u(k-\tau_m)|^{(2p+1)} \right) - \sum_{d=1}^{D} a_d y(k-\tau_d) \qquad (5.2)$$

Where c_p are the polynomial coefficients, b_m are the FIR filter coefficients, a_d are the IIR filter coefficients, M is the delay index, and P is the polynomial order. In this study two Hammerstein DPD have been considered. One based on a FIR filter for modeling the linear block (Hammerstein-FIR) and another on an IIR filter for modeling the linear block (only poles – compare Fig. 5.11, Hammerstein-IIR), as it is shown in (5.3) and (5.4), respectively.

$$y(k) = \sum_{m=0}^{M-1} b_m \left(\sum_{p=1}^{P} c_p u(k-\tau_m)|u(k-\tau_m)|^{2(p-1)} \right) \qquad (5.3)$$

$$y(k) = \sum_{p=0}^{P} c_p u(k-\tau_p)|u(k-\tau_p)|^{2(p-1)} - \sum_{d=1}^{D} a_d y(k-\tau_d) \tag{5.4}$$

In order to choose the best delays contributing in the Hammerstein DPD (FIR and IIR), a preceding study was carried out. It consisted in selecting the most significant past samples of the input ($\tau_1 \ldots \tau_m$) and of the output ($\tau_1 \ldots \tau_D$), among a subset of delays considered for the search, that minimize the estimation error. For this purpose, a FOM was chosen (NMSE) and several model validations with different delays configuration were done. Further details on this study can be found in [3].

Moreover, Hammerstein DPD can be implemented in a digital signal processor using look-up tables (LUTs), as it is shown in Fig. 5.12, where MSD means most significant delays.

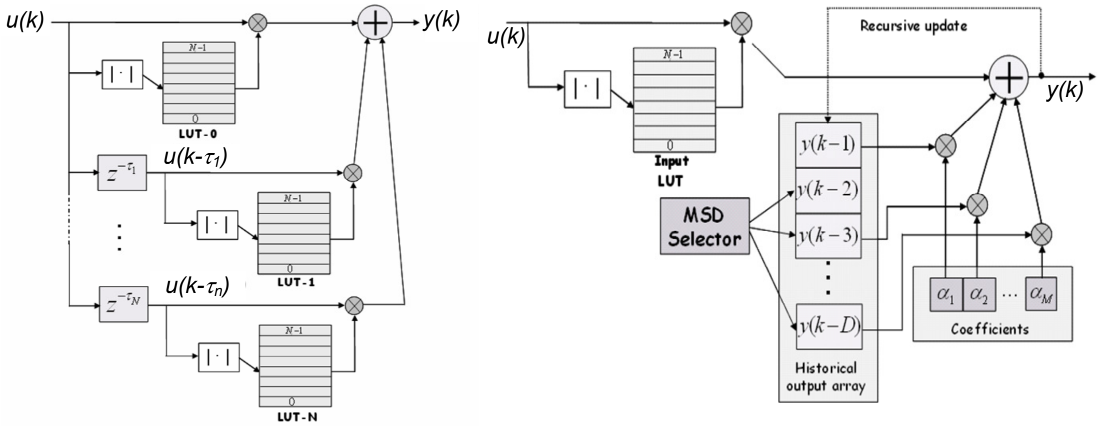

Figure 5.12: *Hammerstein-FIR and IIR LUTs implementation.*

5.4.4 Simulation Results

A simulation was carried out to avail the performance and viability of the Hammerstein pre-distorter. Fig. 5.13 shows simulation pre-distorted output power spectra using memoryless DPD, IIR and FIR Hammerstein DPD.

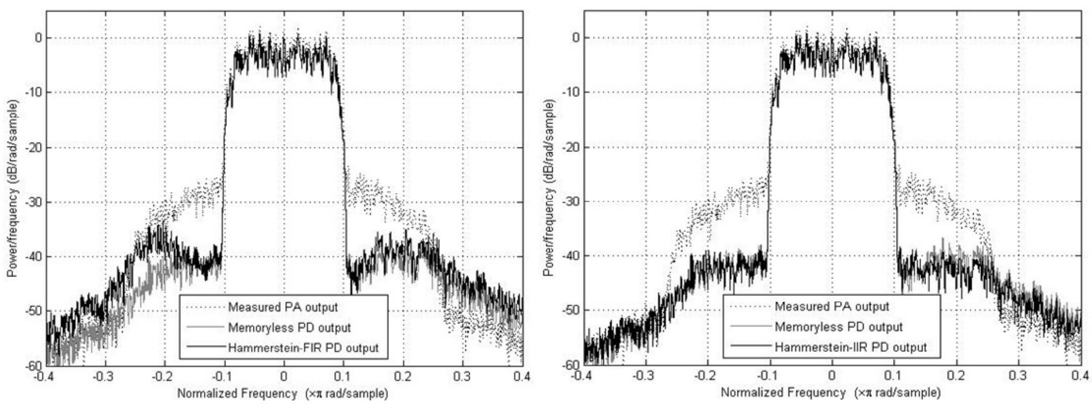

Figure 5.13: *Simulated output power spectra Hammerstein FIR and IIR PD (2 dB IBO).*

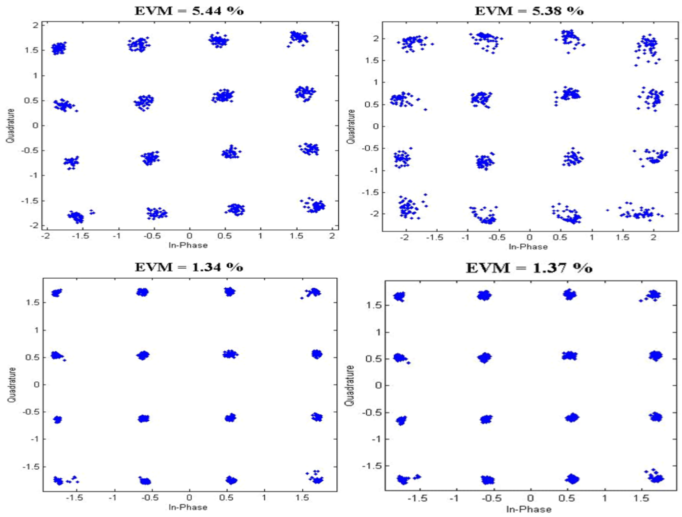

Figure 5.14: *Simulated 16-QAM constellation for: a) Measured PA data (EVM=5.44%), b) Memoryless PD (EVM=5.38%), c) Hammerstein FIR PD (EVM=1.34%), d) Hammerstein IIR PD (EVM=1.37%).*

Around −10 dB of ACPR improvement was achieved using both FIR and IIR Hammerstein DPD. Note that even memoryless DPD can obtain linearity improvement, although very discrete (correction in constellation rotation). In addition, regarding the in-band compensation, Fig. 5.14 shows at least 4% EVM reduction when using both Hammerstein DPD, although this cannot be accomplished when a memoryless pre-distorter is used.

Simulation results present significant improvement in both in-band and out-of-band distortion compensation by using Hammerstein DPD. These results will be compared in the next subsection with experimental results.

5.4.5 Experimental Results

The measurement system is shown in Fig. 5.3, where the modulated input signal is first pre-distorted at baseband (S1 open and S2 closed), up-converted and fed into the PA. Results on the ACPR and EVM reduction achieved by these pre-distorters are listed in Table 5.4. Fig. 5.15 shows the amplified 16-QAM RRC filtered signal output power spectrum, and both output power spectra corresponding to the FIR and IIR Hammerstein DPD.

Table 5.4: *Results on pre-distorters compensation performance.*

	Lower ACPR (dBr)	**Upper ACPR (dBr)**	**EVM (%)**
PA	−32.5	−32.0	7.5
Hammerstein-FIR	−37.0	−37.0	5.0
Hammerstein-IIR	−37.5	−35.5	3.5

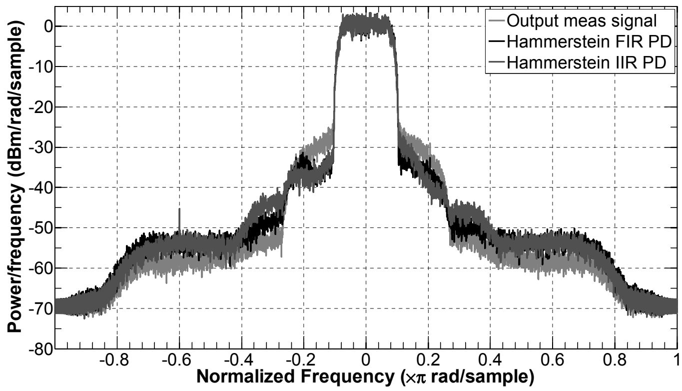

Figure 5.15: *PA and Hammerstein DPD FIR and IIR output power spectra.*

5.4.6 Conclusion

Practical results show a reasonable DPD performance in comparison with simulations, some spectral regrowth is compensated and in-band distortion is significantly improved. ACPR reduction depends on the AM/AM nonlinear compensation, rather than on memory effects compensation. Table 5.4 shows that since power spectrum regrowth is approximately symmetric, nonlinear effects are dominant in front of possible memory effects, so memoryless DPD could compensate such out of band emission. But, regarding in-band distortion, which can be assumed to be shown in the demodulated constellation and measured in terms of EVM, memory effects need to be compensated in order to avoid distortion. Both experimental and simulation results show significant EVM improvement when using Hammerstein DPD. Moreover, the IIR Hammerstein DPD can achieve better linearization results using fewer taps than the FIR Hammerstein DPD. The only drawback regarding the IIR Hammerstein DPD model is the possible unstabilities due to the feedback introduced by the IIR filter. Therefore, future work will be focused in obtaining better nonlinear dynamic models, that consider dynamic interactions of nonlinearities, in order to better match the AM/AM and AM/PM compensation, together with an efficient memory effects estimation. The Wiener model was too simple to represent a PA, and other models have to be developed, for better modeling and pre-distortion results. More accurate models, capable of representing linear and nonlinear memory, could improve the identification results, leading to even more precise linearizes. These models are VS approximations, studied in the next section.

5.5 A Behavioral Model Based on the Wiener-Bose Structure

In this section, a strategy to extract a first-zone filtered behavioral model based on Wiener-Bose structure is introduced, and improvements applied to this model identification are highlighted.

5.5.1 Introduction

The model described in this section is an approximation of VS for complex baseband input/output signals, using linear estimation techniques. Its derivation and particularities are explained. A closed form equation for determining the number of independent parameters for the Wiener-Bose model parametrization process is described. A method to improve the Hessian condition number in the model calculation procedure is presented. The modeling and validation results based on measurements of an industry standard input signal (WCDMA) are displayed. It is shown that this model is capable to represent efficiently PA memory effects with good accuracy.

5.5.2 Measurement Setup

The PA modeled (amplifier A) is described in section 5.2.1. The measured input/output signals were obtained from the measurement setup described in section 5.2.2 (WCDMA signal).

5.5.3 Model Extraction

In this subsection, sub-matrices corresponding to each model order of the regression matrix are explained: $\mathbf{U}_1$, $\mathbf{U}_3$, $\mathbf{U}_5$, $\mathbf{U}_7$. These sub-matrices are latter grouped into one $\mathbf{U}$ matrix:

$$\mathbf{U} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_3 & \mathbf{U}_5 & \mathbf{U}_7 \end{bmatrix} \tag{5.5}$$

Then the model parametrization process is described, and the final model is presented.

Polynomial Used

The selected basis function was an odd-order baseband polynomial [27]. This structure was applied to form different time delay terms of the regression matrices. The polynomial is constructed as (4.27), repeated here for convenience:

$$\phi_p[\mathbf{u}(k)] = \sum_{p=1}^{P} c_p |\mathbf{u}(k)|^{2(p-1)} \mathbf{u}(k) \tag{5.6}$$

First Order

The first order kernel matrix was based on the ordinary FIR regression matrix [40]. It is a $(N - M + 1) \times M$ matrix, where N represents the number of elements of the input vector used in the model extraction process and M represents the number of taps considered in the filter or the memory of the system.

$$U_1 = \begin{bmatrix} \mathbf{u}(k) & \mathbf{u}(k-1) & \ldots & \mathbf{u}(k-M+1) \end{bmatrix} \tag{5.7}$$

Third Order

The third order kernel structure describes every third order interaction between the instantaneous and the delayed version of the input. It is a $(N - M + 1) \times M^3$ matrix, before profiting from the symmetry properties of the Volterra kernels, and $(N - M + 1) \times [M^2(M + 1)/2]$ after that. A single example with $M = 2$ resulting in 8 elements is shown in (5.8).

$$\begin{aligned}\mathbf{U}_3 = [&\mathbf{u}(k)\mathbf{u}(k)\mathbf{u}^*(k) \quad \mathbf{u}(k)\mathbf{u}(k)\mathbf{u}^*(k-1) \quad \mathbf{u}(k)\mathbf{u}(k-1)\mathbf{u}^*(k) \\ &\mathbf{u}(k)\mathbf{u}(k-1)\mathbf{u}^*(k-1) \quad \mathbf{u}(k-1)\mathbf{u}(k)\mathbf{u}^*(k) \quad \mathbf{u}(k-1)\mathbf{u}(k)\mathbf{u}^*(k-1) \\ &\mathbf{u}(k-1)\mathbf{u}(k-1)\mathbf{u}^*(k) \quad \mathbf{u}(k-1)\mathbf{u}(k-1)\mathbf{u}^*(k-1)]\end{aligned} \tag{5.8}$$

From this example, it is possible to conclude that the 5^{th} and the 6^{th} elements of the matrix can be discarded in the estimation process, because they are exactly the same as the 3^{rd} and the 4^{th} elements (symmetry). This results in a reduced rank matrix.

Fifth and Seventh Orders

The fifth and seventh order kernels are every fifth and seventh order interactions between the input and its delayed versions.

The fifth order matrix is $(N - M + 1) \times M^5$ (full) and $(N - M + 1) \times [M^2(M + 1)^2(M + 2)/12]$ (symmetric).

The seventh order matrix is $(N - M + 1) \times M^7$ (full) and $(N - M + 1) \times [M^2(M + 1)^2(M + 2)^2(M + 3)/144]$ (symmetric).

Model Parametrization

The Wiener-Bose model is linear in its parameters and can be solved using linear regression methods as Least-Squares (LS).

The final Hessian has the structure as:

$$\mathbf{H} = \mathbf{U}^{\mathrm{H}}\mathbf{U} = \begin{bmatrix} \mathbf{U}_1^*\mathbf{U}_1 & \mathbf{U}_1^*\mathbf{U}_3 & \mathbf{U}_1^*\mathbf{U}_5 & \mathbf{U}_1^*\mathbf{U}_7 \\ \mathbf{U}_3^*\mathbf{U}_1 & \mathbf{U}_3^*\mathbf{U}_3 & \mathbf{U}_3^*\mathbf{U}_5 & \mathbf{U}_3^*\mathbf{U}_7 \\ \mathbf{U}_5^*\mathbf{U}_1 & \mathbf{U}_5^*\mathbf{U}_3 & \mathbf{U}_5^*\mathbf{U}_5 & \mathbf{U}_5^*\mathbf{U}_7 \\ \mathbf{U}_7^*\mathbf{U}_1 & \mathbf{U}_7^*\mathbf{U}_3 & \mathbf{U}_7^*\mathbf{U}_5 & \mathbf{U}_7^*\mathbf{U}_7 \end{bmatrix} \tag{5.9}$$

So interactions between all matrices are considered in the Hessian.

For very large matrices, the Strassen inversion method [51] was applied to reduce the rank of the matrices to be inverted, and the ordinary LS solution was used.

Final Model

Fig. 5.16 shows the final structure used for the parametrization of the model with the regression sub-matrices.

The resulting parameter vector of the LS solution contains the Wiener-Bose coeffi-

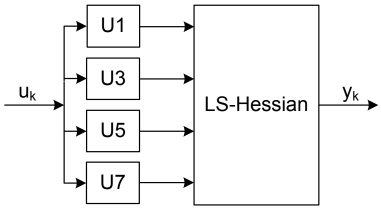

Figure 5.16: *Final Wiener-Bose model structure with regression sub-matrices.*

cients and has the length:

$$\begin{aligned} L = \overbrace{M}^{1^{st}order} + \overbrace{M^2\frac{(M+1)}{2}}^{3^{rd}order} + \\ + \overbrace{M^2\frac{(M+1)^2(M+2)}{12}}^{5^{th}order} + \\ + \underbrace{M^2\frac{(M+1)^2(M+2)^2(M+3)}{144}}_{7^{th}order} \end{aligned} \tag{5.10}$$

From this structure it is possible to identify the sequentially organized coefficients corresponding to the kernels.

5.5.4 Improving the Condition Number of the Ill-Conditioned Hessian

A problem when estimating high order kernels is that the LS Hessian, (5.9), can be singular due to the products $\mathbf{U}_5^{\mathrm{H}}\mathbf{U}_i$ and $\mathbf{U}_7^{\mathrm{H}}\mathbf{U}_i$ (for $i = 1, 3, 5, 7$). If the columns in the matrices $\mathbf{U}_5$ and $\mathbf{U}_7$ present very small differences between themselves, a very high condition number (CN) will arise due to the almost linear dependency.

This effect happens because the sample rate used to acquire the signal was so high that the input was not a persistent excitation for these nonlinearities (5^{th} and 7^{th} order). On the other hand, a lower sample rate would not allow to capture the desired 7^{th} IMD order using this measurement setup. So, although the excitation signal had a strong high order inner correlation in short-periods of time, it may be less correlated when observed in longer periods.

To the best of the author's knowledge, other works in the area of baseband Volterra-based behavioral models do not consider this phenomenon (e. g., [36,70]). To avoid this effect, the delays considered were multiplied by an integer resampling factor rf, and consequently the time between each sample was increased as in (5.11). This technique reduces the CN and can lead to a positive definite Hessian if rf is chosen sufficiently high (typically 2 or 3), this way improving the estimation process.

$$\mathbf{U}_1 = \begin{bmatrix} \mathbf{u}(k.rf) & \mathbf{u}[(k-1).rf] & \ldots & \mathbf{u}[(k-M+1).rf] \end{bmatrix} \tag{5.11}$$

Table 5.5 shows the CNs for models of different orders using 3 delay taps with

$rf = 1$ to 5. The CN in bold represent that the Hessian was singular. In these cases, an increase of rf was necessary to find the solution.

It is important to note that this process needs to be applied to all matrices involved in the estimation process ($\mathbf{U}_1$, $\mathbf{U}_3$, $\mathbf{U}_5$ and $\mathbf{U}_7$), and not only to the ones that became singular ($\mathbf{U}_5$ and $\mathbf{U}_7$).

Table 5.5: *Hessian CN for models of* 3^{rd}*,* 5^{th} *and* 7^{th} *order*

rf	**1**	**2**	**3**	**4**	**5**
H_3	$5.9e9$	$1.3e7$	$7.4e5$	$1.9e4$	$1.0e3$
H_5	**1.2e18**	$2.6e12$	$5.1e9$	$3.2e7$	$3.0e5$
H_7	**3.4e18**	**3.3e18**	$1.1e14$	$1.5e11$	$2.5e8$

5.5.5 Validation

Fig. 5.17 shows AM/AM-conversion limiting curves of the measured output signal (reference) and of the estimated models using 2 and 4 delay taps. These two figures highlight an important characteristic of VS that the Wiener-Bose model also has, i. e., a higher order and number of coefficients improve not only the NMSE of the models, but also the model capacity to effectively represent amplifier memory effects. This property cannot be observed if the NMSE or the PSD are used as the only metrics (see Table 5.6 and Fig. 5.18). The higher order models showed a good accuracy in

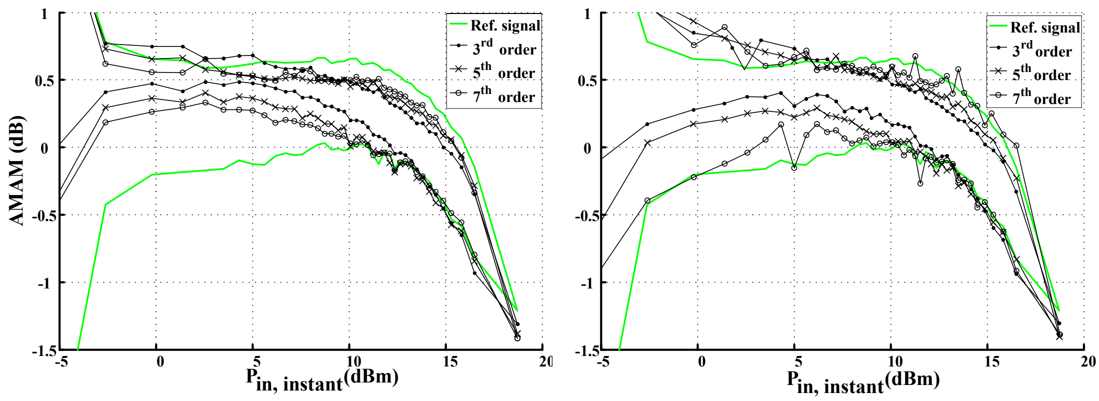

Figure 5.17: *AM/AM-conversion limiting curves of the reference and models of* 3*rd,* 5*th, and* 7*th order using* 2*(left side) and* 4*(right side) delay taps.*

modeling the linear and especially the nonlinear memory, what confirms that this kind of memory is strongly present on the signal higher order interactions. Models that do not care about these interactions can not represent nonlinear memory correctly [21].

Table 5.6: *NMSE*(dB) *for models of* 3^{rd}*,* 5^{th} *and* 7^{th} *order*

	3rd order	**5th order**	**7th order**
2 delay taps	-33.0	-33.8	-34.1
4 delay taps	-33.1	-34.2	-34.2

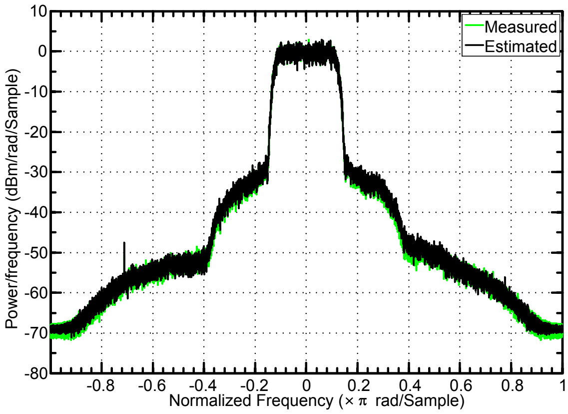

Figure 5.18: *PSD of the measured and modeled output data. The results are from a 7^{th} order model using 4 delay taps.*

5.5.6 Conclusion

This section described the extraction of a new first zone filtered behavioral model using the Wiener-Bose structure, improvements in the estimation process, and its properties.

The well-organized model configuration allows to determine precisely the input signal interactions. An easy identification of the input signal interactions is a very useful property when pruning VS models, as only the most important interactions should be selected among a high number of parameters.

A method to improve the LS Hessian used in the estimation process was introduced. Models of different orders and number of coefficients were estimated and validated from measured input/output data, and good NMSE results were achieved. It was shown that the estimated models were capable of representing linear and nonlinear memory.

Although Wiener-Bose models have good accuracy and the capacity of representing linear and nonlinear memory, they have a high number of parameters, and can eventually have worse results than simpler models due to the noise modeling, as will be shown in next section. New FOMs will be developed and a performance comparison will be established, in order to determine which models are better for different systems presenting distinct types of memory.

5.6 A Figure of Merit for Nonlinear Distortion

This section presents alternatives to correct NMSE drawbacks, explained and shown with examples in subsection 2.4.1.

5.6.1 Introduction

The PA output signal is as a sum of an input correlated part $y_{\mathrm{C}}(k)$ and an input uncorrelated part $y_{\mathrm{N}}(k)$, each composed by a static and a dynamic part

$$\begin{aligned} y(k) &= y_{\mathrm{C}}(k) + y_{\mathrm{N}}(k) \\ &= y_{\mathrm{CS}}(k) + y_{\mathrm{CD}}(k) + y_{\mathrm{NS}}(k) + y_{\mathrm{ND}}(k) \end{aligned} \tag{5.12}$$

The correlated part of the signal $y_{\mathrm{C}}(k)$ is determined by the best linear approximation (BLA), defined as in (3.11) for time domain, and (3.30) for frequency domain [33, 52]. The BLA is usually calculated by the receiver equalizer. The correlated components are the main part of the output signal, not crucial for nonlinear modeling purposes, but they have an enormous weight in the ordinary NMSE leading to erroneous conclusions about the efficiency of models in characterizing nonlinear distortions. As an example, in Fig. 5.4, the $y_{\mathrm{C}}(k)$ is 99.6% of the energy of the signal, and the components of $y_{\mathrm{ND}}(k)$ are only 0.4%.

Figures of Merit are defined to quantify the models inaccuracy, in time or frequency domain. Although mathematically correct, they can heavily weight the correlated part of the signals, as well as synchronization issues, and put little weight to the uncorrelated part, the main objective of nonlinear PA models. In order to correct this behavior, new FOMs were developed.

5.6.2 Mathematical Formulation

The NMSE without the BLA components ($\mathrm{NMSE_{BLA}}$) is found by removing the BLA output components from both measured and modeled signal. This FOM is capable to quantify the amount of the uncorrelated part $y_{\mathrm{N}}(k)$ of the signal taken into account by the model. It is defined as

$$\mathrm{NMSE_{BLA}} = 10\log\left\{\frac{\sum\limits_{k=1}^{N}|y_2^{\mathrm{meas}}(k) - y_2^{\mathrm{mod}}(k)|^2}{\sum\limits_{k=1}^{N}|y_2^{\mathrm{meas}}(k)|^2}\right\} \tag{5.13}$$

Where

$$\begin{aligned} y_2^{\mathrm{meas}}(k) &= y^{\mathrm{meas}}(k) - \mathrm{BLA^{meas}[u(k)]} \\ y_2^{\mathrm{mod}} &= y^{\mathrm{mod}}(k) - \mathrm{BLA^{mod}[u(k)]} \end{aligned} \tag{5.14}$$

The $\mathrm{BLA^{meas}}[u(k)]$ is the BLA estimated between input/output measured signal and $\mathrm{BLA^{mod}}[u(k)]$ is the BLA calculated between the measured input and the model output signal. This calculation is also seen in Fig. 5.19.

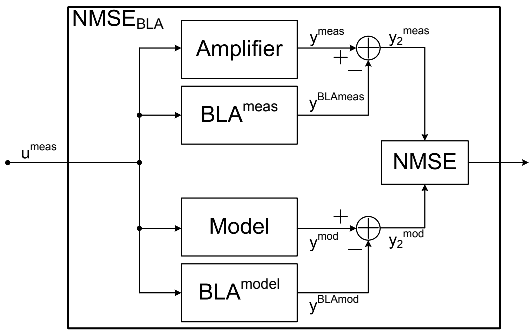

Figure 5.19: *Estimation of the* $\mathrm{NMSE_{BLA}}$.

An example of a measured spectrum signal without BLA output components is given in Fig. 5.20. The distortion is now evident and can be compared, resulting in a clear difference when the NMSE is calculated using the full signal.

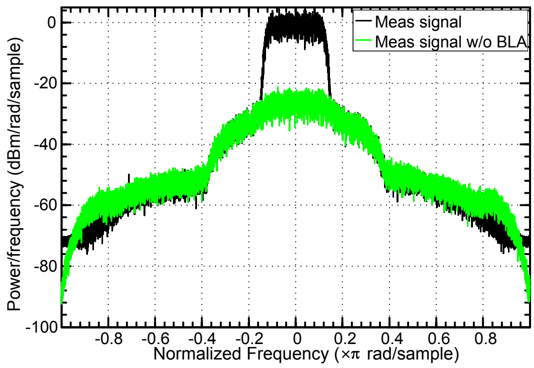

Figure 5.20: *An example of a measured signal without BLA output components.*

Extending this concept, it is possible to define the best nonlinear approximation (BNA) from (5.12)

$$\begin{aligned} y_N(k) = \mathrm{BNA}[\mathrm{u(k)}] &= \mathrm{BNA_S}[\mathrm{u(k)}] + \mathrm{BNA_D}[\mathrm{u(k)}] \\ \mathrm{BNA_S}[\mathrm{u(k)}] &= \sum_p c_{(2p-1)} u(k)|u(k)|^{2(p-1)} \\ \mathrm{BNA_D}[\mathrm{u(k)}] &= y(k) - \mathrm{BNA_S}[\mathrm{u(k)}] \end{aligned} \tag{5.15}$$

It is composed by a static part, $\mathrm{BNA_S}$, and a dynamic part, $\mathrm{BNA_D}$.

The $\mathrm{BNA_S}$ contains the linear and the nonlinear static components of the signal. The coefficients $c^{(p)}$ for the $\mathrm{BNA_S}$ calculation can be estimated performing a LS polynomial fit between the measured input signal and the difference of the measured output signal and the BLA output signal, where p is the maximal order of the measured IMD.

The $\mathrm{BNA_D}[\mathrm{u(k)}]$ is obtained indirectly by the subtraction of the $\mathrm{BNA_S}[\mathrm{u(k)}]$ component from the signal in analysis, and is the objective of behavioral modeling. It contains not only the nonlinear memory, but also other distortion sources included in the signal, like synchronization errors, noise, and estimation errors.

The $\mathrm{NMSE_{BNA_S}}$ is defined as

$$\mathrm{NMSE_{BNA_S}} = 10\log\left\{\frac{\sum\limits_{k=1}^{N}|y_3^{\mathrm{meas}}(k) - y_3^{\mathrm{mod}}(k)|^2}{\sum\limits_{k=1}^{N}|y_3^{\mathrm{meas}}(k)|^2}\right\} \tag{5.16}$$

Where

$$\begin{aligned} y_3^{\mathrm{meas}}(k) &= y^{\mathrm{meas}}(k) - \mathrm{BNA_S^{meas}}[\mathrm{u(k)}] \\ y_3^{\mathrm{mod}} &= y^{\mathrm{mod}}(k) - \mathrm{BNA_S^{mod}}[\mathrm{u(k)}] \end{aligned} \tag{5.17}$$

This FOM suppresses the nonlinear static components of the signal, weighting only the nonlinear distortion. It shows the ability of the model to describe the dynamic effects, including linear and nonlinear memory, and solves the problems of the ordinary NMSE, providing a better insight of the models under test capabilities. Fig. 5.21 shows an spectrum example of a given signal without $\mathrm{BNA_S}$ output components.

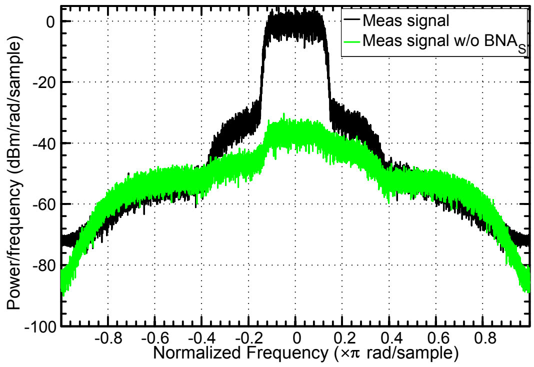

Figure 5.21: *An example of a measured signal without* $\mathrm{BNA_S}$ *output components.*

5.6.3 Models Extraction and Results

The measured input/output signals were obtained from the measurement setup described in section 5.2.2 (WCDMA Signal – Simulation).

From these signals, seven models were estimated for all types of memory effects, and the NMSE, $\mathrm{NMSE_{BLA}}$ and $\mathrm{NMSE_{BNA}}$ were evaluated. The results are shown in Tables 5.7, 5.8, and 5.9, sorted by NMSE results. The abbreviations are: neural networks using 10 neurons with tangent-sigmoid activation functions (NN), seventh order baseband polynomial (BB Poly), look-up table (LUT), parallel Wiener (PW), parallel Hammerstein (PH), and Wiener-Bose (WBose). The PH model has a structure as shown in Fig. 5.22. The models PW, PH, and WBose had a maximum of 3 delay taps. Table 5.10 present these models based on the reference ranking number in Tables 5.7, 5.8, 5.9, and on the FOMs $\mathrm{NMSE_{BLA}}$ and $\mathrm{NMSE_{BNA}}$.

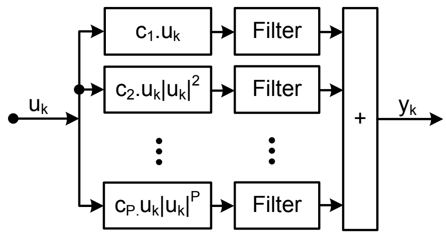

Figure 5.22: *Parallel Hammerstein model.*

Table's Analysis

For the signal from the PA model with linear and nonlinear memory (LM/NLM) in Table 5.10, the ranking is the same as for $\mathrm{NMSE_{BLA}}$. For the $\mathrm{NMSE_{BNA}}$, Saleh model presents better results than BB Poly and NN.

Using the signal from the PA model with linear memory (LM), the LUT had the best overall results, although this is a memoryless model. NN appears as the worst one. BB poly is better than Saleh at $\mathrm{NMSE_{BLA}}$ ranking. For $\mathrm{NMSE_{BNA}}$ ranking, Saleh is the second one.

For the signal from the PA model with nonlinear memory (NLM), WBose showed the best overall results, and LUT comes right after it. Thereafter appears another model with memory, the PH. NN is again the last one. Saleh present better results than BB Poly and PW.

For the signal from the memoryless PA model (noM), LUT present overall best results, while NN again performs worse. The models without memory were better than models with memory in general, except the WBose for NMSE_{BLA}. Saleh stayed on second position for NMSE_{BNA}.

No model presented an overall top performance for the different system memory. The models have to be selected according to the system characteristics, in order to obtain the best performance.

Table 5.7: *NMSE summary – simulated signal LM/NLM and LM*

Model	**LM/NLM**			**LM**		
	NMSE **(dB)**	NMSE BLA **(dB)**	NMSE BNA **(dB)**	NMSE **(dB)**	NMSE BLA **(dB)**	NMSE BNA **(dB)**
WBose	−44.0	−27.0	−13.5	−42.0	−26.5	−11.0
PH	−37.5	−22.0	−7.0	−40.5	−25.0	−9.5
PW	−37.0	−21.5	−7.0	−40.0	−25.0	−9.5
LUT	−35.0	−21.0	−6.5	−38.0	−29.0	−14.0
BB Poly	−34.5	−20.5	−6.0	−36.0	−24.0	−8.5
NN	−32.0	−17.0	−3.0	−31.0	−16.0	−2.0
Saleh	−30.0	−16.5	−6.0	−26.0	−16.0	−12.0

Table 5.8: *NMSE summary – simulated signal NLM*

Model	**NLM**		
	NMSE **(dB)**	NMSE BLA **(dB)**	NMSE BNA **(dB)**
WBose	−44.0	−28.0	−13.0
PH	−37.0	−21.0	−6.5
LUT	−37.0	−21.5	−7.0
PW	−36.0	−21.0	−6.0
BB Poly	−36.0	−6.0	−6.0
NN	−31.0	−15.0	−1.5
Saleh	−30.5	−16.5	−6.5

5.6.4 Conclusion

New FOMs were derived extracting the BLA and the BNA components of the output signals. Using the NMSE_{BLA} and NMSE_{BNA}, a detailed analysis of models' performance was achieved, based on simulated input/output signals.

When systems present linear and nonlinear memory, models should be capable of representing these types of memory effects to accomplish good results. Memoryless

Table 5.9: *NMSE summary – simulated signal noM*

Model	**noM**		
	NMSE **(dB)**	NMSE BLA **(dB)**	NMSE BNA **(dB)**
LUT	−62.0	−49.0	−33.5
WBose	−42.0	−27.0	−11.0
PH	−40.5	−25.0	−9.5
BB Poly	−40.5	−25.0	−9.5
PW	−40.5	−25.0	−9.5
NN	−28.0	−12.5	2.0
Saleh	−25.0	−17.0	−19.0

Table 5.10: *Ranking based on* $\mathrm{NMSE_{BLA}}$ *and* $\mathrm{NMSE_{BNA}}$

LM/NLM		**LM**		**NLM**		**noM**	
NMSE BLA	NMSE BNA	NMSE BLA	NMSE BNA	NMSE BLA	NMSE BNA	NMSE BLA	NMSE BNA
WBose	WBose	LUT	LUT	WBose	WBose	LUT	LUT
PH	PH	WBose	Saleh	LUT	LUT	WBose	Saleh
PW	PW	PH	WBose	PH	PH	PH	WBose
LUT	LUT	PW	PW	BBPoly	Saleh	BBPoly	BBPoly
BBPoly	Saleh	BBPoly	PH	PW	BBPoly	PW	PH
NN	BBPoly	Saleh	BBPoly	Saleh	PW	Saleh	PW
Saleh	NN	NN	NN	NN	NN	NN	NN

models are inefficient in this case.

For systems having only linear memory, representing memory in models has little effect in overall results, as the LUT showed best performance, and the Saleh model was at the second place in $\mathrm{NMSE}_{\mathrm{BNA}}$ ranking. Nonlinearity has a much stronger impact than the linear memory for these systems.

For systems with nonlinear memory, it is necessary to select the correct model structure and achieve good nonlinearity fitting. This characteristics were better achieved by the WBose, but the best nonlinearity fitting was reached by the LUT, which was placed on the second rank. Only after are models as PH and PW, that can represent some amount of nonlinear memory.

Models representing memory may have bad performance for memoryless systems. This fact was presented by the BB Poly which showed a better result than PH and PW.

Saleh had surprisingly good results. This model is known for a good fitting of the nonlinearity with only 4 coefficients. Indeed, fitting memory is important for predistortion purposes and cannot be ignored. But using complex models capable of linear and nonlinear memory estimation as the WBose can deliver worse results than the LUT, for the wrong systems (e.g., memoryless systems). Better methods for fitting memory should be developed, in order to estimate nonlinearity and memory in an efficient way.

NN has the worst results almost in all categories. The NN using only tangent-sigmoid functions was too simple for these signals, and more elaborated (and computationally intensive) ones should be used. These networks are out of the scope of this analysis.

After determining which models have better performance for specific systems, parallel structures combining models which have the best results will be shown in the next section. The proposed models are expected to achieve superior accuracy, allowing the construction of precise pre-distorters.

5.7 An Analysis of Nonlinear Parallel Behavioral Models

This section presents an analysis of actual nonlinear parallel behavioral models, compares their performance, proposes a general pre-processing structure, and also new parallel model configurations.

5.7.1 Introduction

A common strategy to obtain more accurate behavioral models consists in adding several structures in parallel. Parallel models have been successfully employed in nonlinear system identification in different areas using time or frequency-domain data, as in [30, 32, 71]. The advantage of parallel models regards the possibility of integrating several structures in different branches, permitting scalability in the design. The convergence is guaranteed since the next branch is conformed to the residue of the previous one. Some of these models will be analyzed in this section that proposes also a preprocessor for the extraction of parallel PA behavioral models that improves the identification capabilities shown by classical parallel structures. The general pre-processing structure follows the principle of separating the static nonlinear PA behavior in order to allow better identification results. New parallel modeling techniques are introduced, with improved performance in comparison with the actual ones. A comparison of the

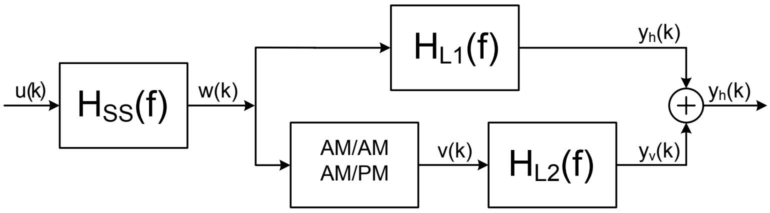

Figure 5.23: *Polyspectral model.*

proposed pre-processing structure with common parallel configurations using linear estimation is presented. Four types of noise-free simulated data obtained from PA models presenting different memory effects were applied, and also an additional measured signal. Finally, three new parallel structures are proposed and the results are described.

5.7.2 Parallel Models

An example of a parallel model that uses time-domain data for its extraction is reported in [30, 71] and shown in Fig. 4.11. The estimation procedure can be summarized as follows: In a first approach, a filter (linear time invariant block) is estimated; then an intermediate signal is obtained as the result of filtering the input signal with the previously obtained filter and finally, a polynomial is estimated by means of this intermediate signal and the output signal. The error generated by this two-step least-squares (LS) estimation is then captured by the next branch of the parallel structure, and so on until no significant improvement in the entire model performance can be noted. This error can be avoided if a suitable estimator for the Wiener structure is used. Neural networks typically apply such nonlinear estimation techniques, but they are a computationally intensive solution. Unfortunately, a linear estimator for a Wiener model would require the same number of parameters as required for a VS, and would not be suitable, but this is the price to pay for generality.

Another example is the modified parallel Hammerstein (PH) [31, 45], that uses one filter for each nonlinear order. Its structure allows determining Volterra kernels in one step. Using exactly the same structure in a second branch does not lead to better results, as the whole dynamic behavior of the PA has been already captured by the first branch. The second branch then will only model noise, and thus its contribution is completely useless. As already reported in [24], parallel models are very sensitive to noise if too many paths are used. Consequently a proper selection of the paths and the order of the nonlinearity should be made to assure low noise and good convergence.

An accuracy improvement in parallel modeling is possible by considering the use of a different structure (than the one used for the first branch) on the second branch, capable of estimating the remains of dynamics that have not been modeled in the first branch. An example of this structure can be found for frequency domain in [32], and it is shown in Fig. 5.23. In these models, a path is designed for the linear part of the signal and another path for the nonlinear one. Treating the residue with another structure, better results were achieved in comparison with a model that presents branches with the same structure. However, the continuation of this process is crucial to obtain better modeling results, as it will be shown in following sections.

5.7.3 A Pre-Processor for PA Linear Estimation

The linear LS estimation combines very interesting characteristics and is a widely employed estimator. Nevertheless, it treats linear and nonlinear components of the data in the same way, and to correct this problem applying weighted LS for every particular input signal is a difficult task. Focusing in the particular case of PA modeling, the linear components present much more power than the nonlinear ones, normally the distortion is at least 30 dB below the carrier. If models are estimated directly from input/output sets of data, both linear and nonlinear behavior will be modeled together and thus the linear part of the signal will appear as noise for the nonlinear part within the estimation. This unwanted situation can be avoided by removing the linear part of the signal. This could be accomplished fitting a best linear approximation between input/output data, in an approach similar to the one described for polyspectral models [32].

Therefore, we now go a step further, and propose a parallel model consisting in a first (upper) pre-processing branch, that represents the memoryless nonlinearity (NL), and subsequent branches, responsible for modeling the remain of the output signal that has not been identified by the first branch. The block diagram is depicted in Fig. 5.24, and an example of the residue for the next estimation using given input/output data records is displayed in Fig. 5.25. The upper branch (pre-processing) removes all the static nonlinearity of the signal, or the noise for the identification of nonlinear distortion parameters, allowing a more accurate identification of the dynamic behavior (in lower branches). The PA dynamic model is estimated between $u_{\text{meas}}(k)$ and $y_{\text{res}}(k)$. The output signal is composed of the sum of $y_{\text{NL}}(k)$ and $y_{\text{dyn}}(k)$, the nonlinear static block and the PA dynamic model output, respectively.

A similar approach was suggested in [30], but limited to build the nonlinearity with the two-tone AM/AM – AM/PM response. Alike in [32], it is possible to replace the memoryless nonlinearity by a filter (upper branch) in order to remove linear PA dynamics to proceed with the identification of the nonlinear part in lower branches. However, the identification performance achieved with this solution is too dependent on the number of coefficients used to describe the filter and thus it looses generality in the comparison. Also in [33] the procedures to extend the model were not given, as pointed out in [21].

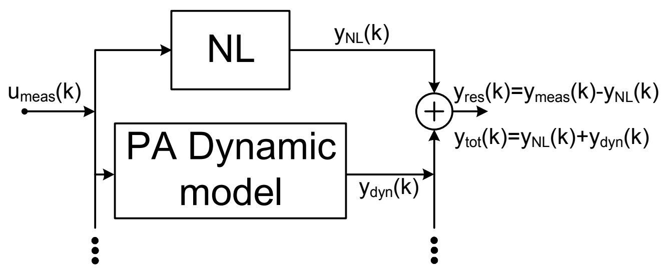

Figure 5.24: *Pre-processing for a dynamic PA model.*

In the following section, several structures for PA modeling (with and without pre-processing) will be compared in terms of normalized mean squared error (NMSE), in order to compare the identification performance achieved when using this pre-processing technique.

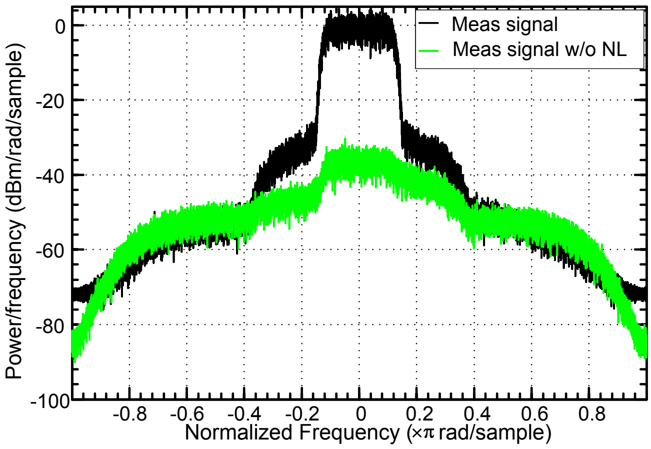

Figure 5.25: *Residue at the output of the NL block (Meas signal w/o NL).*

5.7.4 Models Extraction and Results

Five sets of data were used for model extraction and validation. Four sets were the ones also used in section 5.2.2 (WCDMA – Simulation).

A fifth measured set of data was obtained from a PA with modified Bias-Tee (amplifier B) described in section 5.2.1, using the Wimax signal described in section 5.2.2 (Wimax Signal). The AM/AM and AM/PM characteristics of this modified amplifier are displayed in Fig. 5.26. The measurement setup covered intermodulation distortions up to 7^{th} order. The input/output PSD are shown in Fig. 5.27.

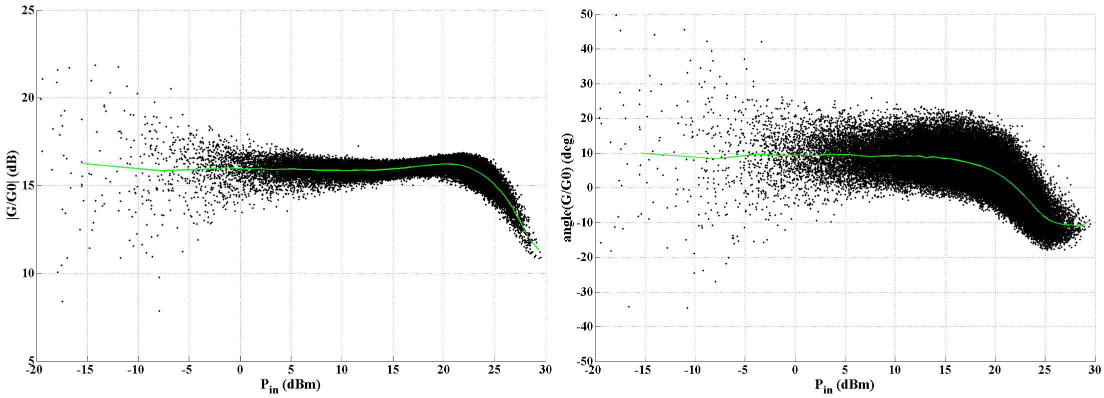

Figure 5.26: *AM/AM and AM/PM characteristics of the modified Bias-Tee amplifier measured signal used for modeling.*

Table 5.11 shows a comparison among the following behavioral models: ordinary baseband power series [26], Wiener, parallel Wiener (PW), and modified parallel Hammerstein models (PH). The figure of merit used to characterize the models' accuracy is the NMSE [32]. Models were extracted considering a 7^{th} order baseband polynomial and a maximum of 3 delay taps. Observing results in Table 5.11, we can see that the parallel Wiener with 2 branches shows practically the same NMSE as a single Wiener model for all different type of data. No significant accuracy improvement, measured in terms of NMSE, is appreciated. The inclusion of additional branches in the model would be useless, since it would be modeling noise and thus not contributing to the

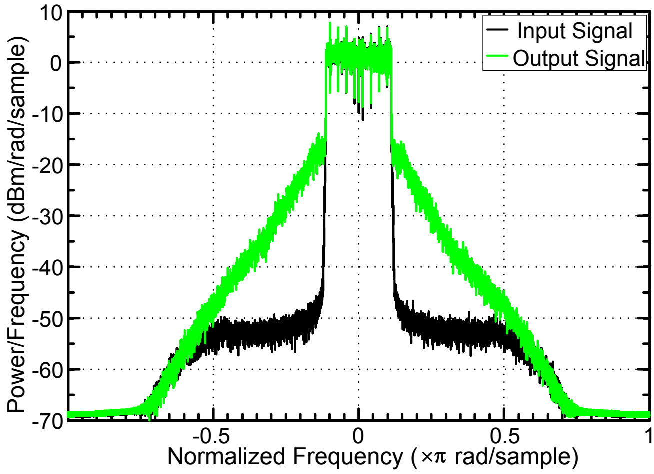

Figure 5.27: *PSD of the modified Bias-Tee amplifier driven by a Wimax input signal.*

final estimation results. The PH model shows the best NMSE results, since Volterra coefficients were estimated in one step. Practically no residue is left for a further branch using the same PH structure. Attending to these results one may conclude that the addition of parallel replicas of the same model structure do not contribute significantly to obtain a more accurate identification performance.

Table 5.11: *NMSE summary – simulated and measured signals*

Model	LM/NLM (dB)	LM (dB)	NLM (dB)	noM (dB)	Meas (dB)
Poly	−34.5	−36.0	−36.0	−40.5	−21.0
Wiener	−37.0	−40.0	−36.0	−40.0	−21.0
PW	−37.0	−40.5	−36.0	−40.5	−21.0
PH	−37.5	−40.5	−37.0	−40.5	−22.0

In order to highlight the advantages of the general pre-processing technique presented in Fig. 5.24, we particularize the general dynamic PA model with a reduced Volterra series (WBose) structure, as it shown in Fig. 5.28. The WBose block follows the Wiener-Bose approach, reported in [17]. For this model only three delay taps have been considered. Moreover, in order to reduce the complexity of the model, the pruning technique proposed in [37] has been used. It was originally derived until 5^{th} order baseband Volterra series (VS), and was extended to the 7^{th} order in this work, as presented in Appendix A.

This behavioral model with pre-processing is composed of a nonlinear memoryless block, implemented with a look-up table (LUT) to avoid the dependency on the polynomial order, and the WBose block.

The estimation procedure for the extraction of the WBose model with pre-processing can be summarized as follows (see Fig. 5.24):

- First, extract the memoryless nonlinear function by means of the measured input ($u_{meas}(k)$) and output ($y_{meas}(k)$) data;
- Then, calculate the residual signal $y_{res}(k)$ defined as the difference between the

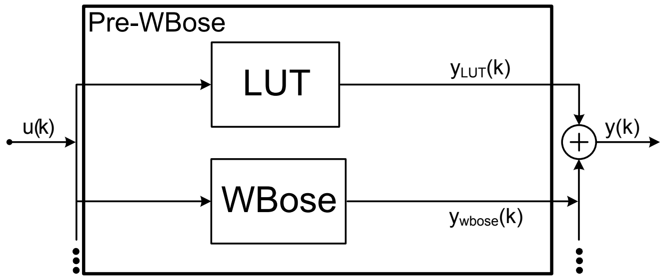

Figure 5.28: *Proposed initial configuration for estimation of PA behavioral models.*

measured output $y_{meas}(k)$ and the output of the memoryless nonlinear model $y_{NL}(k)$;

- Parameterize the dynamic model using the input signal $u_{meas}(k)$ and the residual signal $y_{res}(k)$.

The output of the overall model is the sum of the outputs of both branches, as it is depicted in Fig. 5.28.

Table 5.12 shows the identification results for a memoryless nonlinearity (LUT alone), a PH with pre-processing, a WBose model without pre-processing (WBose), and a WBose model with pre-processing (Pre-WBose).

Table 5.12: *NMSE summary – simulated and measured signal*

Model	LM/NLM (dB)	LM (dB)	NLM (dB)	noM (dB)	Meas (dB)
LUT	−35.0	−38.0	−37.0	−60.0	−21.0
Pre-PH	−38.0	−54.0	−38.0	−64.0	−22.0
WBose	−42.0	−42.0	−42.0	−42.0	−23.0
Pre-WBose	−44.0	−56.0	−44.0	−65.0	−23.0

Unlike the use of parallel replicas (see results in Table 5.11), the use of the proposed pre-processing technique improves the identification results achieved by a single dynamic model without any kind of pre-processing, for all systems presenting memory effects. Therefore, this technique offers the possibility to increase the identification accuracy with only a slight change in the estimation procedure. For signals obtained from memoryless systems, the pre-processing technique with LUT shows improvements of more than 15 dB. Results also show that improvements for this modeling technique using WBose and LUTs are at the limit. Different structures at the subsequent branches could improve the NMSE results. In the following, the data used to compare and validate the proposed architectures and techniques will be the measured data obtained from the modified Bias-Tee amplifier.

5.7.5 Estimating Memory Effects With Sub-Band Structures

A RF power amplifier has a complex structure, presenting many kinds of memory. They can be classified as [72]:

- Low frequency (kHz to MHz): Thermal effects, trapping effects, biasing circuits, AGC loops;

- High frequency (GHz): Transistor (transit time and reactance parasitics), matching networks (group delay).

These memories are mixed together in a PA (nonlinear coupled), and the problem of estimating behavioral models becomes very difficult [73]. Models capable of identifying memory at different signal rates can improve the identification performance. This can be accomplished using parallel sub-band filtering techniques. It is a powerful method to design very large order FIR filters, operating at a high speed, with smaller filters, operating at slower speed, reducing computational complexity, as shown in [74, 75]. In this technique, the input signal is decimated into different rates, filtered for each branch, and later interpolated with the respective branch rate to form the output signal. With smaller filters, the matrix to be inverted in LS estimation process is also smaller. An example of baseband digital pre-distortion using this technique if found in [75]. A system identification problem, where sub-band neural networks were employed to recover audio signals is shown in [76]. It was concluded that sub-band adaptive filters have better performance for highly correlated input signals (also the case for amplifier identification) than full-band adaptive filters. Also in [76] it was proven that, for sub-band input signals, the eigenvalue spread of the sub-band signal will be smaller or equal the eigenvalue spread of the full signal, what guarantees a better condition number of the Hessian matrix used in the LS estimation.

A proposed model to estimate different kinds of memories is presented in Fig. 5.29, with two variants. They will differ in the way in which the resampling factor (rf), already explained in [17], is optimized. The first model has the first branch with rf of 1, thus doing a "blind" first estimation, and then optimizing the other branches considering the other resampling rates. The second model begins the rate optimization already in the first branch. This factor is adjusted for each branch, following a nonlinear search. This is accomplished by varying rf and performing an estimation for this point, recording the results and retaining the best rf for this branch. Another similar work using decimation for optimal estimation was reported in [77], where a way to optimize this resampling factor for ill-conditioned LS Hessian is shown, also with evaluations of the estimated function. When using this resampling factor, only the decimation was employed for the estimation proposed. The interpolation operation cited by references [74–76] was not necessary.

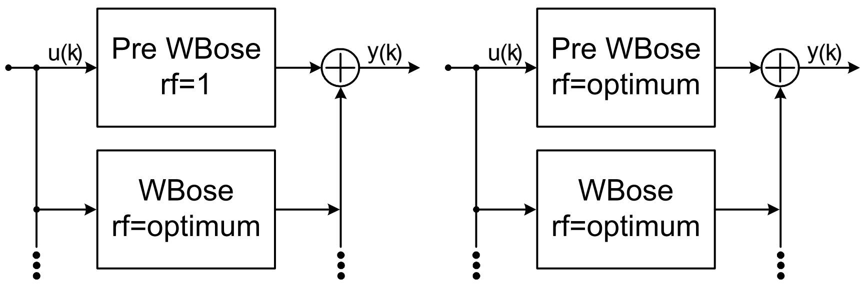

Figure 5.29: *Proposed configuration for estimation of PA behavioral models.*

This degree of freedom for rf can be optimized for "short-term" memory effects ($rf = 1$ to $rf \approx 50$) in single steps, as it has been done in this work. Therefore, the model will be composed of branches optimized for different rates, having improved identification capabilities.

Finally, the overall model output is obtained by the sum of all responses at each branch. The single branches structures (LUT-WBose) are rich enough to capture residues of the signal that still are not modeled.

5.7.6 Extraction of Sub-Band Parallel Models

The models described in Fig. 5.29 were estimated with the same input/output signals as in section 5.7.4, using 5 branches each.

Figure 5.30 shows the results of the *rf* optimization curves for branches from 2 to 5 and considering $rf = 1$ in the first branch.

Figure 5.31 shows the NMSE dependency on *rf* values at the first branch of the model having *rf* optimized for all branches in Fig. 5.29 (right structure), and for remaining branches. No significant changes were noted after $rf > 50$, so *rf* iterations were limited up to 50.

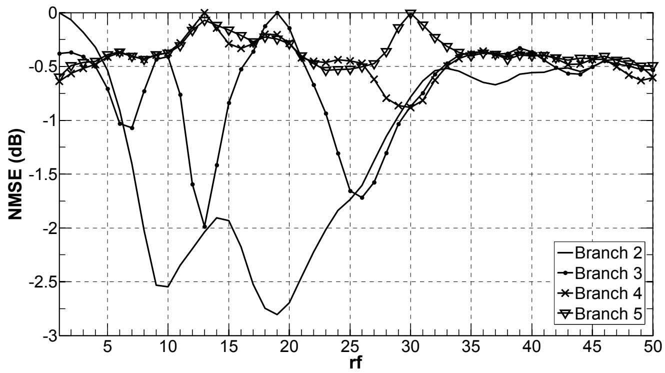

Figure 5.30: *Results for optimization at remaining branches. The best resampling factor were: 19, 13, 30, 3 for branches 2 to 5, respectively.*

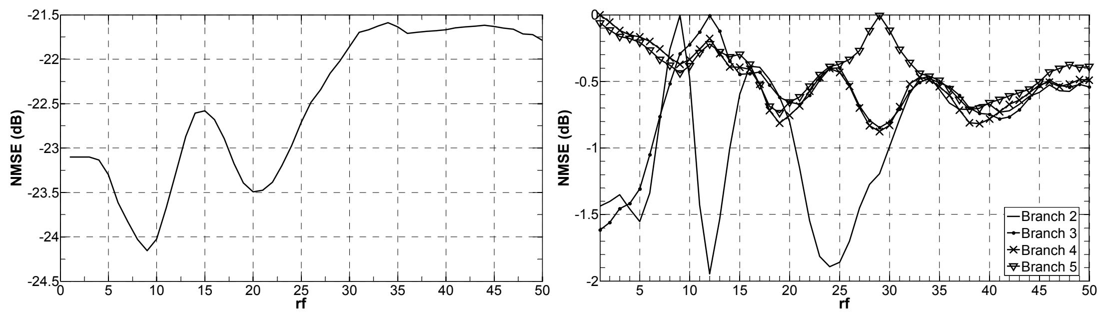

Figure 5.31: *Results for optimization of the first branch (left) and at remaining branches (right). The best resampling factor was found at position 9 for the first branch, and 12, 4, 39, 30 for branches 2 to 5, respectively.*

A third variant of model optimization was tested, using the heuristic search algorithm named Simulated Annealing (SA). This algorithm is used to find the best sparse delays contributing at each branch to the identification of the behavioral model of the PA, as displayed in Fig. 5.32. This nonlinear search technique has been used in the extraction of baseband behavioral models [78]. It searches for the best configuration of delays to improve the identification accuracy. The results achieved in terms of NMSE using this method are also listed in Table 5.13. It can be observed that already in the second branch, the technique with SA parameterize a considerable part of the residue, having a faster convergence than the other methods. However, the total NMSE figure is slightly worse than the other configuration using the rf technique. Nevertheless, the NMSE figure is improved in comparison to the use of a single branch for the identification. The absence of the fifth branch is due to the noise modeling of this branch, that does not contribute to the final results. Another possibility for the initial conditions, having the first branch with higher identification FOM (-24.0 dB NMSE) was also tested (compare Table 5.13), but results were very close to this previous model using SA and will not be reported here.

Table 5.13 presents the NMSE obtained by models at each branch, and their corresponding resampling factor. At the end, the total NMSE achieved by each model, as the result of the contribution of all branches forming the model, is also displayed. Although the model having an optimized rf at the first branch presented a better NMSE figure in the initial branch, the final optimization results were practically the same as the one starting with $rf = 1$. So, initial conditions (or the rf in the first branch) were considered good for both cases. What is significant is that the use of the resampling factor together with parallel models has improved final results by approximately 6 dB NMSE in comparison with results shown in Table 5.12 (results obtained using the measured signal). Although the models presented here have an increased complexity in comparison with previous models, the accuracy is clearly better.

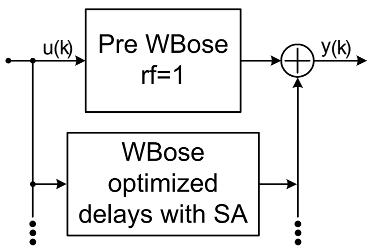

Figure 5.32: *Variation with Simulated Annealing optimization for the subsequent branches.*

Table 5.13: *NMSE, rf and optimal delays results*

First branch $rf = 1$			Optim. for all Branches		Optim. with SA	
Branch	rf	**NMSE (dB)**	rf	**NMSE (dB)**	Delays	**NMSE (dB)**
1	1	−23.0	9	−24.0	[1 2 3]	−23.0
2	19	−3.0	12	−2.0	[1 8 24]	−4.0
3	13	−1.0	4	−1.0	[1 8 6]	−0.5
4	30	−1.0	39	−1.0	[1 20 29]	−0.5
5	1	−1.0	30	−1.0	—	—
Model		−29.0		−29.0		−28.0

Finally, the model that takes into account optimized rf for all branches (Pre-WBose par) was also tested considering all different sets of simulated data. The obtained results were compared with the other modeling techniques presented in Table 5.12

and repeated here for convenience in Table 5.14. For memoryless and linear memory systems, no additional branches were necessary, and results were better than −55 dB NMSE. For systems presenting nonlinear memory, additional branches were necessary to improve final results, staying at −48 dB NMSE. An unmodeled residue was initially present, minimized by subsequent branches. Based on the previous results, the VS pruning technique derived in [37] is more suitable for systems presenting linear memory, fortunately the most often case.

Also improvements comparing with other modeling techniques were evident: 4 dB NMSE for simulated signals, and 6 dB NMSE for measured signals, when compared with Pre-WBose (without parallel additional branches). Comparing with a memoryless model (LUT), improvements were from 5 dB to 18 dB NMSE.

Sub-band parallel models are a suitable possibility to estimate linear memory and also nonlinear memory, throughout estimation residues and sub-band techniques, leading to more accurate results. In order to highlight the spectral improvements of this

Table 5.14: *Comparison NMSE summary – simulated and measured signal*

Model	LM/NLM (dB)	LM (dB)	NLM (dB)	noM (dB)	Meas (dB)
LUT	−35.0	−38.0	−37.0	−60.0	−21.0
Pre-WBose	−44.0	−56.0	−44.0	−65.0	−23.0
Pre-WBose par	−48.0	−56.0	−48.0	−65.0	−29.0
Branches	3	1	4	1	5

technique, Fig. 5.33 shows the measured output signal and the residues of the model with pre-processing and sub-sampling parallel branches (Pre-Wbose par) and the model with only pre-processing (Pre-Wbose) – the NMSE results are listed in Table 5.14, in column *Meas*. The in-band residue improvements of the model employing sub-sampling techniques are clearly seen in this figure.

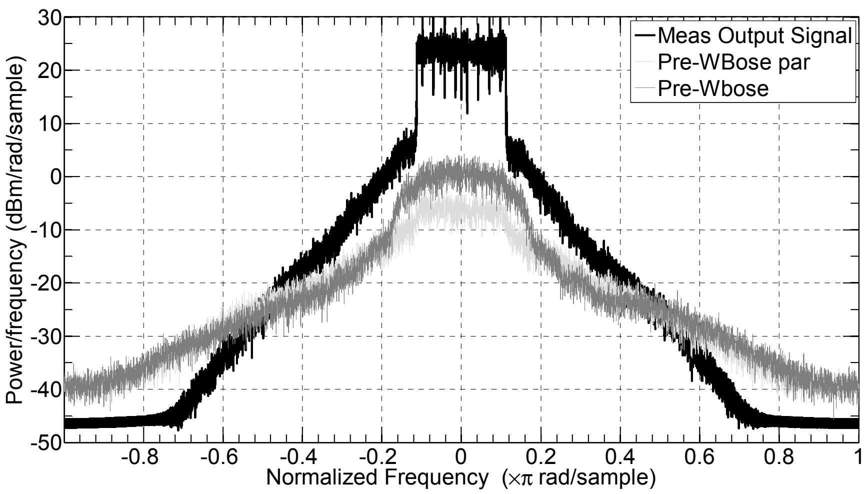

Figure 5.33: *PSD of the measured output signal and of the residue of the model with pre-processing and sub-sampling parallel branches (Pre-Wbose par) and only with pre-processing (Pre-Wbose).*

5.7.7 Conclusion

A general pre-processing technique for PA behavioral modeling has been presented. This technique has shown to be efficient and has been validated using a PH model and a particular configuration based in a reduced Wiener-Bose dynamic PA behavioral model. Later on, three types of parallel models using pre-processing technique in the first branch and reduced VS in the remaining ones were introduced. They employed resampling factor or Simulated Annealing techniques and showed an improved identification performance, in terms of NMSE, in comparison with other simpler models. The main drawback is the increase in the number of parameters used for modeling when considering these techniques. Results have shown that parallel models with different structures using sub-sampling in their subsequent branches can improve the identification performance and justify the inclusion of additional branches. Results have confirmed that the accuracy of a PA behavioral model considering only one single branch is lower than considering pre-processing and resampling techniques. The importance of sub-band parallel models was shown in terms of overall NMSE improvement, and showing an efficient reduction of the residue in the final estimation process.

Chapter 6

Conclusions

This thesis analyzed different methodologies and possibilities to implement BMs, ranging from linear systems, as FIR filters, to accurate PA models, composed by parallel cascades of look-up tables and reduced order VS approximation.

A general overview of BMs was initially presented, and main difficulties in estimating a PA BM were highlighted. The most used FOMs were analyzed for their strengths and weaknesses. A reliable FOM is necessary, since it offers a possibility to quantify the model's performance. Parametric FOMs are very suitable for use in algorithms, but lack from information about memory effects contained in nonparametric (graphical) FOMs. In general, more than one FOM is necessary for a complete analysis. Improved alternatives for the NMSE were proposed ($\mathrm{NMSE_{BLA}}$ and $\mathrm{NMSE_{BNA}}$), for a better evaluation of the modeled nonlinear distortions components of the signal.

Linear estimation techniques for linear systems were surveyed, and a de-noising estimator for linear systems was proposed, based on SVD, that reorganized the filter taps sequence. This technique improves the system memory estimation.

Then the PA was analyzed as a nonlinear system. Memoryless models were studied, including orthogonal polynomials for complex-valued input signals and LUTs. The latter presented the best overall results, employed after in other estimation techniques. An analysis of basic nonlinear systems with memory was performed. Wiener, Hammerstein, and their equivalent parallel models were explored in details. Their particularities relative to linear estimation were identified, e. g., Wiener models cannot be estimated directly using linear estimation techniques, and Hammerstein models include only the main diagonals of the Volterra kernels. They are very simplified models, which do not account for dynamic interaction of nonlinearities, insufficient for an amplifier characterization operating near compression, the main interest of this study.

More elaborated models were studied and developed, capable of VS approximations. The Wiener-Bose (WBose) model was adapted for the complex-valued input/output signal case, its regression matrix was well defined, and alternatives to improve the CN of the Hessian matrix were established. Further on, pruning techniques were applied to the WBose model, making it very suitable for amplifier estimation, due to the lower number of parameters in comparison with a full VS. Good results and suitable characteristics encouraged further studies with the pruned WBose model. It was integrated in a parallel configuration, after a pre-processing technique which improved the PA estimation. The pre-processing consisted in removing the static nonlinearities of the signal, so only the dynamic part was to be identified by the model. The linear part of the signal, which is much bigger than the nonlinear distortions (the main interest of BM), was then removed. This technique, combined with the use of a resampling factor for pruned WBose models in a parallel configuration, showed to be very efficient for PA modeling. The distortions caused by the PA's different memories (thermal effects, trapping effects, biasing circuits, AGC loops, transistor transit time and reactance parasitics, matching networks) can then be better identified, so more accurate pre-distorters can be developed.

Further studies are the improvement of Wiener-Bose models, with other pruning techniques (e. g., based on the input/output signal), and the full automatization of this process. Also of interest are other parallel models, as well as the implementation and analysis of their performance in pre-distortion systems. On-line adaption routines can be implemented in a fast and efficient way for computers, as all models are based on linear estimation techniques.

Appendix A

Derivation of Volterra Kernels Reduction up to 7^{th} Order

This appendix present an extension to the 7^{th} order from the pruning technique presented in [37], to be used with (4.3). The extension was necessary due to the possibility to obtain more accurate models, and for comparisons with other modeling strategies.

Table A.1: *Equations* 3^{th}*order*

1)	$x(n)x(n)x(n)$	$h_3(0,0,0)$
2)	$x(n)x^2(n-i)$	$h_3(0,i,i)$

Table A.2: *Term 2*

$h_3(0,i,i)$	$\tilde{h}_3(0^*,i,i)$
	$\tilde{h}_3(0,i^*,i)$

Table A.3: *Equations* 5^{th}*order*

1)	$x(n)x(n)x(n)x(n)x(n)$	$h_5(0,0,0,0,0)$
2)	$x(n)x(n)x(n)x^2(n-i)$	$h_5(0,0,0,i,i)$
3)	$x(n)x(n)x^3(n-i)$	$h_5(0,0,i,i,i)$
4)	$x(n)x^4(n-i)$	$h_5(0,i,i,i,i)$
5)	$x(n)x^2(n-i)x^2(n-i_2)$	$h_5(0,i_1,i_1,i_2,i_2)$
6)	$x^2(n-i)x^3(n-i_2)$	$h_5(i_1,i_1,i_2,i_2,i_2)$

Table A.4: *Terms 1 to 4*

$h_5(0,i,i,i,i)$	$\tilde{h}_5(0^*,i^*,i,i,i)$
	$\tilde{h}_5(0,i^*,i^*,i,i)$

Table A.5: *Term 5*

$h_5(0,i_1,i_1,i_2,i_2)$	$\tilde{h}_5(0^*,i_1^*,i_1,i_2,i_2)$
	$\tilde{h}_5(0^*,i_1,i_1,i_2^*,i_2)$
	$\tilde{h}_5(0,i_1^*,i_1^*,i_2,i_2)$
	$\tilde{h}_5(0,i_1^*,i_1,i_2^*,i_2)$
	$\tilde{h}_5(0,i_1,i_1,i_2^*,i_2^*)$

Table A.6: *Term 6*

$h_5(i_1,i_1,i_2,i_2,i_2)$	$\tilde{h}_5(i_1^*,i_1^*,i_2,i_2,i_2)$
	$\tilde{h}_5(i_1^*,i_1,i_2^*,i_2,i_2)$
	$\tilde{h}_5(i_1,i_1,i_2^*,i_2^*,i_2)$

Table A.7: *Equations 7^{th}order*

1)	$x(n)x(n)x(n)x(n)x(n)x(n)x(n)$	$h_7(0,0,0,0,0,0,0)$
2)	$x(n)x(n)x(n)x(n)x(n)x^2(n-i)$	$h_7(0,0,0,0,0,i,i)$
3)	$x(n)x(n)x(n)x(n)x^3(n-i)$	$h_7(0,0,0,0,i,i,i)$
4)	$x(n)x(n)x(n)x^4(n-i)$	$h_7(0,0,0,i,i,i,i)$
5)	$x(n)x(n)x^5(n-i)$	$h_7(0,0,i,i,i,i,i)$
6)	$x(n)x^6(n-i)$	$h_7(0,i,i,i,i,i,i)$
A	$x(n)x(n)x(n)x^2(n-i)x^2(n-i_2)$	$h_7(0,0,0,i_1,i_1,i_2,i_2)$
B	$x(n)x(n)x^2(n-i)x^3(n-i_2)$	$h_7(0,0,i_1,i_1,i_2,i_2,i_2)$
C	$x(n)x^3(n-i)x^3(n-i_2)$	$h_7(0,i_i,i_1,i_1,i_2,i_2,i_2)$
D	$x(n)x^2(n-i_1)x^2(n-i_2)x^2(n-i_3)$	$h_7(0,i_1,i_1,i_2,i_2,i_3,i_3)$
E	$x(n)x^2(n-i_1)x^4(n-i_2)$	$h_7(0,i_1,i_1,i_2,i_2,i_2,i_2)$
F	$x^2(n-i_1)x^5(n-i_2)$	$h_7(i_1,i_1,i_2,i_2,i_2,i_2,i_2)$
G	$x^2(n-i_1)x^2(n-i_2)x^3(n-i_3)$	$h_7(i_1,i_1,i_2,i_2,i_3,i_3,i_3)$
H	$x^3(n-i_1)x^4(n-i_2)$	$h_7(i_1,i_1,i_1,i_2,i_2,i_2,i_2)$

Table A.8: *Terms 1 to 6*

$h_7(0,i,i,i,i,i,i)$	$\tilde{h}_7(0^*,i^*,i^*,i,i,i,i)$
	$\tilde{h}_7(0,i^*,i^*,i^*,i,i,i)$

Table A.9: *Term A*

$h_7(0,0,0,i_1,i_1,i_2,i_2)$	$\tilde{h}_7(0^*,0^*,0^*,i_1,i_1,i_2,i_2)$
	$\tilde{h}_7(0,0^*,0^*,i_1^*,i_1,i_2,i_2)$
	$\tilde{h}_7(0,0^*,0^*,i_1,i_1,i_2^*,i_2)$
	$\tilde{h}_7(0,0,0^*,i_1^*,i_1^*,i_2,i_2)$
	$\tilde{h}_7(0,0,0^*,i_1^*,i_1,i_2^*,i_2)$
	$\tilde{h}_7(0,0,0^*,i_1,i_1,i_2^*,i_2^*)$
	$\tilde{h}_7(0,0,0,i_1^*,i_1^*,i_2^*,i_2)$
	$\tilde{h}_7(0,0,0,i_1,i_1^*,i_2^*,i_2^*)$

Table A.10: *Term B*

$h_7(0,0,i_1,i_1,i_2,i_2,i_2)$	$\tilde{h}_7(0^*,0^*,i_1^*,i_1,i_2,i_2,i_2)$
	$\tilde{h}_7(0^*,0^*,i_1,i_1,i_2^*,i_2,i_2)$
	$\tilde{h}_7(0,0^*,i_1^*,i_1^*,i_2,i_2,i_2)$
	$\tilde{h}_7(0,0^*,i_1^*,i_1,i_2^*,i_2,i_2)$
	$\tilde{h}_7(0,0,i_1^*,i_1^*,i_2^*,i_2,i_2)$
	$\tilde{h}_7(0,0,i_1,i_1^*,i_2^*,i_2^*,i_2)$
	$\tilde{h}_7(0,0,i_1,i_1,i_2^*,i_2^*,i_2^*)$

Table A.11: *Term C*

$h_7(0,i_1,i_1,i_1,i_2,i_2,i_2)$	$\tilde{h}_7(0^*,i_1^*,i_1^*,i_1,i_2,i_2,i_2)$
	$\tilde{h}_7(0^*,i_1^*,i_1,i_1,i_2^*,i_2,i_2)$
	$\tilde{h}_7(0^*,i_1,i_1,i_1,i_2^*,i_2^*,i_2)$
	$\tilde{h}_7(0,i_1^*,i_1^*,i_1^*,i_2,i_2,i_2)$
	$\tilde{h}_7(0,i_1,i_1^*,i_1^*,i_2^*,i_2,i_2)$
	$\tilde{h}_7(0,i_1,i_1,i_1^*,i_2^*,i_2^*,i_2)$
	$\tilde{h}_7(0,i_1,i_1,i_1,i_2^*,i_2^*,i_2^*)$

Table A.12: *Term D*

$h_7(0, i_1, i_1, i_2, i_2, i_3, i_3)$	$\tilde{h}_7(0^*, i_1^*, i_1^*, i_2, i_2, i_3, i_3)$
	$\tilde{h}_7(0^*, i_1^*, i_1, i_2^*, i_2, i_3, i_3)$
	$\tilde{h}_7(0^*, i_1^*, i_1, i_2, i_2, i_3^*, i_3)$
	$\tilde{h}_7(0^*, i_1, i_1, i_2^*, i_2^*, i_3, i_3)$
	$\tilde{h}_7(0^*, i_1, i_1, i_2, i_2, i_3^*, i_3^*)$
	$\tilde{h}_7(0, i_1^*, i_1^*, i_2^*, i_2, i_3, i_3)$
	$\tilde{h}_7(0, i_1^*, i_1^*, i_2, i_2, i_3^*, i_3)$
	$\tilde{h}_7(0, i_1^*, i_1, i_2^*, i_2^*, i_3, i_3)$
	$\tilde{h}_7(0, i_1^*, i_1, i_2^*, i_2, i_3^*, i_3)$
	$\tilde{h}_7(0, i_1^*, i_1, i_2, i_2, i_3^*, i_3^*)$
	$\tilde{h}_7(0, i_1, i_1, i_2^*, i_2^*, i_3^*, i_3)$
	$\tilde{h}_7(0, i_1, i_1, i_2, i_2^*, i_3^*, i_3^*)$

Table A.13: *Term E*

$h_7(0, i_1, i_1, i_2, i_2, i_2, i_2)$	$\tilde{h}_7(0^*, i_1^*, i_1^*, i_2, i_2, i_2, i_2)$
	$\tilde{h}_7(0^*, i_1^*, i_1, i_2^*, i_2, i_2, i_2)$
	$\tilde{h}_7(0^*, i_1, i_1, i_2^*, i_2^*, i_2, i_2)$
	$\tilde{h}_7(0, i_1^*, i_1^*, i_2^*, i_2, i_2, i_2)$
	$\tilde{h}_7(0, i_1, i_1^*, i_2^*, i_2^*, i_2, i_2)$
	$\tilde{h}_7(0, i_1, i_1, i_2^*, i_2^*, i_2^*, i_2)$

Table A.14: *Term F*

$h_7(i_1, i_1, i_2, i_2, i_2, i_2, i_2)$	$\tilde{h}_7(i_1^*, i_1^*, i_2^*, i_2, i_2, i_2, i_2)$
	$\tilde{h}_7(i_1, i_1^*, i_2^*, i_2^*, i_2, i_2, i_2)$
	$\tilde{h}_7(i_1, i_1, i_2^*, i_2^*, i_2^*, i_2, i_2)$

Table A.15: *Term G*

$h_7(i_1, i_1, i_2, i_2, i_3, i_3, i_3)$	$\tilde{h}_7(i_1^*, i_1^*, i_2^*, i_2, i_3, i_3, i_3)$
	$\tilde{h}_7(i_1^*, i_1^*, i_2, i_2, i_3^*, i_3, i_3)$
	$\tilde{h}_7(i_1, i_1^*, i_2^*, i_2^*, i_3, i_3, i_3)$
	$\tilde{h}_7(i_1, i_1^*, i_2, i_2, i_3^*, i_3^*, i_3)$
	$\tilde{h}_7(i_1, i_1^*, i_2^*, i_2, i_3^*, i_3, i_3)$
	$\tilde{h}_7(i_1, i_1, i_2^*, i_2^*, i_3^*, i_3, i_3)$
	$\tilde{h}_7(i_1, i_1, i_2^*, i_2, i_3^*, i_3^*, i_3)$
	$\tilde{h}_7(i_1, i_1, i_2, i_2, i_3^*, i_3^*, i_3^*)$

Table A.16: *Term H*

$h_7(i_1, i_1, i_1, i_2, i_2, i_2, i_2)$	$\tilde{h}_7(i_1^*, i_1^*, i_1^*, i_2, i_2, i_2, i_2)$
	$\tilde{h}_7(i_1^*, i_1^*, i_1, i_2^*, i_2, i_2, i_2)$
	$\tilde{h}_7(i_1, i_1, i_1^*, i_2^*, i_2^*, i_2, i_2)$
	$\tilde{h}_7(i_1, i_1, i_1, i_2^*, i_2^*, i_2^*, i_2)$

Bibliography

[1] P. B. Kenington, *High-Linearity RF Amplifier Design*, Artech House, 2000.

[2] J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*, Artech House, 2003.

[3] P. L. Gilabert, G. Montoro, and E. Bertran, “A methodology to model and predistort short-term memory nonlinearities in power amplifiers”, *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*, vol. 1, pp. 142–145, 2006.

[4] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, 1999.

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

[6] M. O’Droma and A. A. Goacher, “Deliverable D 1.3.2.7 final report on linearisation evaluation map”, Tech. Rep., TARGET, 2005.

[7] D. Zhou and V. E. DeBrunner, “Novel adaptive nonlinear predistorters based on the direct learning algorithm”, *IEEE Transactions on Signal Processing*, vol. 55, pp. 120 – 133, 2007.

[8] H. W. Kang, Y. S. Cho, and D. H. Youn, “On compensating nonlinear distortions of an ofdm system using an efficient adaptive predistorter”, *IEEE Transactions on Communications*, vol. 47, pp. 522 – 526, 1999.

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

[10] G. Montoro, P. L. Gilabert, E. Bertran, A. Cesari, and D. D. Silveira, “A new digital predictive predistorter for behavioral power amplifier linearization”, *IEEE Microwave and Wireless Components Letters*, vol. 17, pp. 448–450, June 2007.

[11] J. Wood and D. E. Root, *Fundamentals of Nonlinear Behavioral Modeling of RF and Microwave Design*, Artech House, 2005.

[12] D. T. Westwick and R. E. Kearney, “Identification of physiological systems using pseudo-inverse based deconvolution”, *IEEE-EMBC and CMBEC, Theme 6: Physiological Systems/Modelling and Identification*, vol. 6, pp. 1405–1406, 1995.

[13] D. Silveira, M. Gadringer, M. Mayer, and G. Magerl, "Analysis of RF-power amplifier modeling performance using a 16-QAM modulation over AWGN channels", *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*, vol. 1, pp. 164–167, 2006.

[14] D. Silveira, M. Gadringer, H. Arthaber, M. Mayer, and G. Magerl, "Modeling, analysis and classification of a PA based on identified Volterra kernels", *13th GAAS Symposium*, vol. 1, pp. 405–408, 2005.

[15] D. Silveira, M. Gadringer, P. Gilabert, G. Montoro, Y. Lei, E. Srinidhi, E. Lima, V. Camarchia, M. Pirola, G. Magerl, E. Bertran, A. Goacher, M. ODroma, G. Kompa, and S. Donati, "Comparison of RF power amplifier behavioural models estimated from shared measurement data", *Target Days 2006, Book of Proceedings*, vol. 1, no. ISBN 3-902477-07-5, pp. 129–132, 2006.

[16] P. L. Gilabert, D. D. Silveira, G. Montoro, and G. Magerl, "RF-power amplifier modeling and predistortion based on a modular approach", *The 1st European Microwave Integrated Circuits Conference*, vol. 1, pp. 265–268, 2006.

[17] D. D. Silveira and G. Magerl, "Extraction and improvements of a behavioral model based on the Wiener-Bose structure used for baseband Volterra kernels estimation", *International Microwave Symposium*, vol. 1, pp. 2007–2010, June 2007.

[18] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, Krieger Publishing Company, 1980.

[19] M. Schetzen, "Nonlinear system modeling based on the wiener theory", *Proceedings of the IEEE*, vol. 69, no. 12, pp. 1557–1573, 1981.

[20] W. J. Rugh, *Nonlinear System Theory - The Volterra/Wiener Approach*, Johns Hopkins University Press, 1981.

[21] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches", *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 1150–1163, 2005.

[22] J. Sjoberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P. Glorennec, H. Hjalmarsson, and A. Juditsky, "Nonlinear black-box modeling in system identification: a unified overview", *Automatica*, vol. 31, pp. 1691–1724, 1995.

[23] D. T. Westwick and R. E. Kearney, *Identification of Nonlinear Physiological Systems*, IEEE Press, 2003.

[24] V. Z. Marmarelis, *Nonlinear Dynamic Modeling of Physiological Systems*, J. Wiley & Sons, 2004.

[25] R. K. Pearson, "Selecting nonlinear model structures for computer control", *Journal of Process Control*, vol. 13, pp. 1–26, 2003.

[26] S. Benedetto, E. Biglieri, and R. Daffara, "Modeling and performance evaluation of nonlinear satellite links – a Volterra series approach", *IEEE Transactions Aerospace Electronic Systems*, vol. 15, pp. 494–506, July 1979.

[27] S. Benedetto and E. Biglieri, *Principles of Digital Transmission*, Kluwer Academic, 1999.

[28] R. Raich and G. T. Zhou, "Orthogonal polynomial for complex Gaussian processes", *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2788–2797, October 2004.

[29] R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomials for power amplifier modeling and predistorter design", *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 1468–1479, 2004.

[30] H. Ku, M. D. McKinley, and J. S. Kenney, "Quantifying memory effects in RF power amplifiers", *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 2843–2849, 2002.

[31] H. Ku and J. Stevenson Kenney, "Behavioral modeling of nonlinear RF power amplifiers considering memory effects", *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 2495–2504, 2003.

[32] C. P. Silva, A. A. Moulthrop, and M. S. Muha, "Introduction to polyspectral modeling and compensation techniques for wideband communications systems", *Automatic RF Techniques Group Conference Digest*, vol. 40, pp. 1–15, 2001.

[33] C. P. Silva, C. J. Clark, A. A. Moulthrop, and M. S. Muha, "Survey of characterization techniques for nonlinear communication components and systems", *IEEE Aerospace Conference*, vol. 1, pp. 1–25, 2005.

[34] P. L. Gilabert, G. Montoro, and A. Cesari, "A recursive digital predistorter for linearizing RF power amplifiers with memory effects", *Asia Pacific Microwave Conference*, vol. 2, pp. 1043–1047, 2006.

[35] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, Wiley Interscience, 2000.

[36] A. Zhu, M. Wren, and T. J. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers", *IEEE MTT Symposium Digest*, vol. 1, pp. 787–790, 2003.

[37] A. Zhu, J. C. Pedro, and T. R. Cunha, "Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge", *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, pp. 813–821, 2007.

[38] L. E. Scales, *Introduction to Nonlinear Optimization*, MacMillan Publisher Ltda, 1985.

[39] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice Hall, 2000.

[40] O. Nelles, *Nonlinear System Identification*, Springer, 2001.

[41] S. Ablameyko, L. Goras, M. Gory, and V. Piuri, *Neural Networks for Instrumentation, Measurement and Related Industrial Applications*, vol. 185, NATO Science Series, 2003.

[42] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*, Artech House, 2000.

[43] E. Aschbacher, *Digital Predistortion of Microwave Power Amplifiers*, PhD thesis, Vienna University of Technology, 2005.

[44] A. Hagenblad and L. Ljung, "Maximum likelihood identification of Wiener models with a linear regression initialization", *IEEE Conference on Decision and Control*, vol. 1, pp. 712–713, December 1998.

[45] M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers", *IEEE Transactions on Microwave Theory and Techniques*, vol. 1, pp. 348–359, January 2006.

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

[47] M. S. Muha, C.J. Clark, A.A. Moulthrop, and C.P. Silva, "Validation of power amplifier nonlinear block models", *IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 759–762, 1999.

[48] *ETSI, Digital Cellular Telecommunications System (Phase 2+); Radio Transmission and Reception, 3GPP TS 45.005 version 6.9.0 Release 6, April 2005.*

[49] Mathworks, "Matlab version 7", January 2005.

[50] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, 1993.

[51] M. Abramowitz and A. I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1965.

[52] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 2002.

[53] M. Lortie and R. E. Kearney, "Robust identification of time-varying system dynamics with non-white inputs and outputs noise", *Proceedings of* 20^{th} *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 20, no. 6, pp. 3036–3037, 1998.

[54] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, 1987.

[55] M. C. Jeruchim, P. Balaban, and K. Sam Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*, Kluwer, 2000.

[56] B. Boashash, E. J. Powers, and A. M. Zoubir, *Higher Order Statistical Signal Processing*, Longman House, 1995.

[57] D. Lichtblau and E. W. Weisstein, "Condition Number", *From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/ConditionNumber.html*, December 2004.

[58] A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers", *IEEE Transactions on Communications*, vol. 29, pp. 1715–1720, 1981.

[59] H. Chen, "Modeling and identification of parallel nonlinear systems: Structural classification and parameter estimation methods", *Proceedings of the IEEE*, vol. 83, no. 1, pp. 39–66, January 1995.

[60] T. M. Panicker and V. J. Mathews, "Parallel-cascade realizations and approximations of truncated Volterra systems", *IEEE Transactions on Signal Processing*, vol. 46, pp. 2829 – 2832, 1996.

[61] M. J. Korenberg, "Parallel cascade identification and kernel estimation for nonlinear systems", *Annals of Biomedical Engineering*, vol. 19, pp. 429–455, 1991.

[62] S. Boyd and L. Chua, "Fading memory and the problem of approximating nonlinear operators with Volterra series", *IEEE Transactions on Circuits and Systems*, vol. 32, no. 11, pp. 1150–1161, Nov 1985.

[63] A. Zhu and T. J. Brazil, "Rf power amplifier behavioral modeling using volterra expansion with laguerre functions", *IEEE MTT Symposium Digest*, 2005.

[64] J. G. Proakis, *Digital Communications*, McGraw Hill, 2002.

[65] *IEEE Computer Society and the IEEE Microwave Theory and Techniques Society, 802.16 IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Std 802.16-2004.*

[66] H. AbdulHussein Al Asady and M. Ibnkahla, "Performance evaluation and total degradation of 16-QAM modulations over satellite channels", *CCECE-CCGEI*, vol. 1, pp. 1187–1190, 2004.

[67] A. Conti, D. Dardari, and V. Tralli, "On the performance of CDMA systems with nonlinear amplifier and AWGN", *IEEE Sixth International Symposium on Spread Spectrum Techniques and Applications*, vol. 1, pp. 197–202, 2000.

[68] R. Kumar, D. Taggart, C. Chen, and N. Wagner, "Simulation and modeling of amplifier nonlinearities with 16-QAM modulated waveforms in wireless communication systems", *IEEE 60th Vehicular Technology Conference*, vol. 1, pp. 4212–4216, 2004.

[69] M. ODroma, E. Bertran, M. Gadringer, S. Donati, A. Zhu, P.L. Gilabert, and J. Portilla, "Developments in predistortion and feedforward adaptive power amplifier linearisers", *Gallium Arsenide and Other Semiconductor Application Symposium (GAAS)*, vol. 1, pp. 337 – 340, 2005.

[70] C. Cheng and E. J. Powers, "Optimal Volterra kernel estimation algorithms for a nonlinear communication system for PSK and QAM inputs", *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 147–162, January 2001.

[71] M. J. Korenberg, "Identifying nonlinear difference equation and functional expansion representations: the fast orthoghonal algorithm", *Annals of Biomedical Engineering*, vol. 16, pp. 123–142, 1988.

[72] E. Ngoya and A. Soury, "Modeling memory effects in nonlinear subsystems by dynamic Volterra series", *IEEE Behavioral Modeling and Simulation Workshop*, 2003.

[73] N. Le Gallou, E. Ngoya, H. Buret, D. Barataud, and J. M. Nebus, "An improved behavioral modeling technique for high power amplifiers with memory", *IEEE MTT Symposium Digest*, vol. 1, pp. 983–986, 2001.

[74] S. K. Mitra, A. Mahalonobis, and T. Saramaki, "A generalized structural subband decomposition of FIR filters and its application in efficient FIR filter design and implementation", *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 40, pp. 363–374, 1993.

[75] O. Hammi, S. Boumaiza, M. Jaidane, and F. M. Ghannouchi, "Baseband digital predistortion using subband filtering technique", *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 1699–1702, 2003.

[76] G. Cocchi and A. Uncini, "Subband neural networks prediction for on-line audio signal recovery", *IEEE Transactions on Neural Networks*, vol. 13, pp. 867–876, 2002.

[77] A. Sano and H. Tsuji, "Optimal sampling rate for impulse response identification based on decimation and interpolation", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 460–463, 1993.

[78] P. Gilabert, D. Silveira, G. Montoro, M. Gadringer, and E. Bertran, "Heuristic algorithms for power amplifier behavioral modeling", *IEEE Microwave and Wireless Components Letters*, vol. 17, pp. 715–717, 2007.