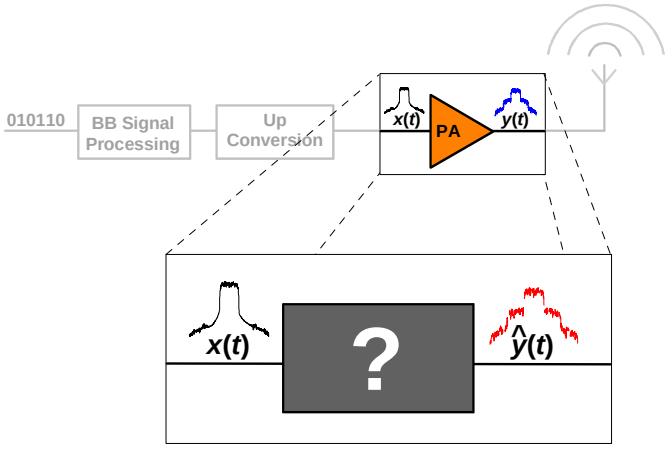
Universidade de Aveiro
2007
Departamento de Electrónica, Telecomunicações e Informática

Pedro Miguel Ribeiro Lavrador

Contribuição ao estudo do impacto das não linearidades nos sistemas de telecomunicações



Universidade de Aveiro
2007
Departamento de Electrónica, Telecomunicações e Informática

Pedro Miguel Ribeiro Lavrador

Contribuição ao estudo do impacto das não linearidades nos sistemas de telecomunicações

Contribution to the study of the impact of nonlinearities on telecommunications systems

tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Engenharia Electrotécnica, realizada sob a orientação científica do Dr. José Carlos Pedro, Professor Catedrático do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro e sob a co-orientação científica do Dr. Nuno Borges Carvalho Professor Associado do mesmo departamento.

Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

Dedico este trabalho à minha esposa Raquel, aos meus pais e irmãs.

o júri

presidente

Doutor Jorge Ribeiro Frade
Professor Catedrático da Universidade de Aveiro (em representação da Reitora da Universidade de Aveiro)

Doutor José Carlos Esteves Duarte Pedro
Professor Catedrático da Universidade de Aveiro (orientador)

Doutor João José Lopes da Costa Freire
Professor Associado com Agregação do Instituto Superior Técnico da Universidade Técnica de Lisboa.

Doutor Tomás António Mendes de Oliveira e Silva
Professor Associado da Universidade de Aveiro

Doutor Nuno Miguel Gonçalves Borges de Carvalho
Professor Associado da Universidade de Aveiro (co-orientador)

Doutor Anding Zhu
Investigador na University College Dublin

agradecimentos

Em primeiro lugar agradeço aos meus orientadores Prof. José Carlos Pedro e Prof. Nuno Borges Carvalho, por todo o apoio que me deram ao longo destes anos e sem o qual este trabalho não teria sido possível. Agradeço o bom ambiente de trabalho proporcionado e o estímulo a procurar sempre a excelência.

Agradeço à Universidade de Aveiro, em particular ao Departamento de Electrónica Telecomunicações e Informática e ao Instituto de Telecomunicações a disponibilização dos recursos materiais necessários à execução do trabalho. Este agradecimento é extensível aos seus colaboradores que sempre estiveram disponíveis para ajudar.

Agradeço à Fundação para a Ciência e Tecnologia o apoio financeiro durante os primeiros anos deste projecto sob a forma de Bolsa de Doutoramento. Agradeço também à Comissão Europeia, na forma da rede de excelência TARGET, o financiamento durante a parte final deste trabalho.
Agradeço à Fundação Luso Americana para o Desenvolvimento o financiamento para viagens de apresentação de artigos em conferências internacionais, apoio este que permitiu o contacto com investigadores internacionais de grande prestígio e o correspondente alargar de horizontes para o meu trabalho.

Agradeço aos colegas de trabalho e amigos, de modo especial ao Pedro Cabral e ao João Paulo Martins companheiros de percurso em grande parte deste trajecto e que tantas vezes me encorajaram a persistir na busca de novas e melhores soluções. A sua amizade e companheirismo foram determinantes ao longo deste trabalho. Agradeço de modo especial o apoio do Pedro Cabral para a redacção desta tese.

Agradeço aos meus pais por terem ajudado a criar em mim desde pequeno o gosto pelo saber e a cultura do procurar saber mais. Agradeço-lhes os sacrifícios que fizeram para me permitir chegar até esta etapa. Sem eles e sem o seu exemplo de esforço nada disto seria possível.

À minha namorada e esposa Raquel, quero agradecer toda a paciência que teve comigo. O suporte nos momentos mais difíceis e o alento para continuar.

A todos o meu Muito Obrigado!

palavras-chave

Modelação Comportamental, Amplificadores de Potência, Série de Volterra

resumo

Esta tese insere-se na área de Electrónica de Rádio Frequência e Microondas e visa o desenvolvimento de ferramentas que permitam a melhor compreensão e análise do impacto da distorção não linear produzida em amplificadores de potência no desempenho de um sistema de telecomunicações sem fios.
Devido à crescente complexidade dos amplificadores a simulação baseada em representações de circuito equivalente tornou-se extremamente pesada do ponto de vista computacional. Assim têm surgido várias técnicas de simulação de sistemas baseadas em modelos comportamentais, ou seja, que tentam aproximar a resposta do sistema a um sinal de entrada, independentemente dos elementos físicos que implementam o circuito.
Neste trabalho foram estudadas as principais técnicas de modelação comportamental existentes assim como as principais características de um amplificador de potência que o modelo comportamental deve ser capaz de prever.
Uma nova formulação de um modelo comportamental baseado na série de Volterra é apresentada em conjunto com o método de extracção ortogonal dos seus coeficientes. A principal vantagem deste novo método de extracção é permitir a determinação independente de cada valor coeficiente na série, garantindo-se deste modo um modelo com uma capacidade de aproximação óptima. A determinação dos coeficientes na série de modo independente é conseguida com base na reorganização dos termos da série e na identificação ortogonal de cada componente de saída.
Adicionalmente, a identificação das componentes de saída de uma não linearidade é ainda utilizada na definição de uma métrica que permite avaliar de modo simples qual é a degradação imposta à qualidade do sinal ao ser passado num amplificador não linear. Esta métrica contabiliza simultaneamente a degradação imposta pelo ruído e pela distorção.

keywords Behavioural Modelling, Power Amplifiers, Volterra Series.

abstract

This thesis is related to the RF and Microwave Electronics field and the main goal of this thesis is to develop tools that can contribute to understand and analyse the impact of nonlinear distortion generated by power amplifiers on wireless communication systems.

Due to the growing complexity of amplifiers, equivalent circuit based simulations become a heavy computational task due to the large number of nonlinear elements to account for. So, several system simulation techniques have been proposed based on behavioural modelling, that is, models that can approximate the system's response to a given input signal regardless of the physical circuit implementation description.

In this thesis, the most important behavioural modelling techniques have been studied as well as the main power amplifier characteristics that the behavioural model should account for.

A new formulation of a Volterra series based behavioural model is presented as well as the corresponding coefficient orthogonal extraction procedure. The main advantage of this new extraction method is to allow the independent determination of the exact value of each coefficient, guaranteeing this way an optimum approximation condition. The exact coefficient determination is achieved by reorganizing the series terms to reach independent subsets and by identifying separately each of systems' output components.

In addition, nonlinearity output component separation is also used to define a Figure of Merit that allows the simple evaluation of signal quality degradation when passed through a nonlinear amplifier. This Figure takes into account simultaneously the impact of noise and distortion.

Table of Contents

List of Figures

List of Acronyms

ADS	Advanced Design System
AM-AM	Amplitude Modulation to Amplitude Modulation
AM-PM	Amplitude Modulation to Phase Modulation
ANN	Artificial Neural Network
BLA	Best Linear Approximator
Bw	Bandwidth
CAD	Computer Aided Design
CDMA	Code Division Multiple Access
CW	Continuous Wave
dB	Decibel
dBm	Decibel related to one milliwatt
DC	Direct Current
FIR	Finite Impulse Response
FoM	Figure of Merit
f_{NL}	Nonlinear Function
GSM	Global System for Mobile communications
H_L	Best Linear Approximator function
IMD	Inter-Modulation Distortion
IMD3	3rd order IMD
IMD5	5th order IMD
IIR	Infinite Impulse Response
IP3	3rd Order Intercept Point
M	Memory Span of the truncated Volterra series
M-QAM	M-ary Quadrature Amplitude Modulation
N	Maximum Order of the truncated Volterra series
NDF	Noise and Distortion Figure
NF	Noise Figure
NFIR	Nonlinear Finite Impulse Response
NIIR	Nonlinear Infinite Impulse Response
NIM	Nonlinear Integral Model
NMSE	Normalized Mean Square Error
PA	Power Amplifier
PAR	Peak to Average Ratio
PF	Polynomial Filter
PSD	Power Spectral Density
RF	Radio Frequency
SINAD	Signal to Noise and Distortion Ratio
SNR	Signal to Noise Ratio
TARGET	Top Amplifier Research Groups in a European Team
UMTS	Universal Mobile Telecommunications System
VS	Volterra Series
WCDMA	Wideband CDMA

1. Introduction

There is some controversy about who was the inventor of wireless radio transmissions. Probably this achievement can not be attributed to a single person but is the result of several contributions. Some of the most important were: the theoretical work developed by Maxwell, the first practical controlled synthesis of radio waves by Hertz, and also, the first practical information transmission over a system by Tesla. There is a US patent of 1900 by Tesla, describing an apparatus with many "*valuable uses as for instance, when it is desired to transmit intelligible messages to great distances […]"*. In Europe Marconi made the first wireless transmission across the English Channel on March 1899. More consensual is that since then we have been assisting, and participating, in one rapidly increasing evolution in the wireless communications.

The development of new wireless communication technologies that occurred in the past few decades was one of the most important revolutions in the last centuries. This development changed the conventional way how people interact with each other. This on-going revolution in communications is self regenerating, as people's eager for new services is fed by the service providers with new products and applications. Both, the number of users and also the services per user have increased considerably. Consequently, the

communication systems need to accommodate the increased information flow generated by the growing consumer's demands.

In both wired and wireless communications the modulation schemes have changed in order to allow higher communication rates with an efficient use of the available spectra. In wired communications, the power consumption and bandwidth used are not major issues when compared to mobile communications. Due to its nature the power is available from 'the line' and the bandwidth is most of the times devoted to the service and not severely interfering with other systems due to its guided propagation nature. In mobile wireless communications, both these aspects might be bottlenecks. The power because the user devices are operated by batteries and the device autonomy is a key factor for its success. The bandwidth is a bottleneck, because the communication channel is shared between all the users and even between different communication systems.

In wireless communication systems the use of the frequency spectra must obey very stringent rules in order to make possible the coexistence of different services and also different service providers. Each communication standard imposes strict spectral masks created to keep the interference between different users and systems at reasonable levels and thus also to guarantee the quality of service. The recent wireless communication protocols (GSM, CDMA2000, UMTS, etc) make use of complex modulation techniques that intend to maximize the data throughput of the communication channel. To increase the debit through the channel, linear transmission should be guaranteed, so that the signal perturbations caused by distortion are avoided. This issue is even more serious when multilevel modulation techniques like M-QAM are used. As they have non-constant amplitude envelopes and the distortion impact might be different for each amplitude level.

The three restrictions referred: (i) the limited power available on the device, (ii) the efficient use of spectra and (iii) the complex modulation schemes used lead to a critical design compromise on the mobile handset power amplifier (PA). Effectively, in order to transmit the modulated signal while avoiding the transmission/generation of spectral garbage the PA should operate on its linear regime. However, to operate the PA on the linear regime an output power back off of several dBs is required, which compromises the PA efficiency and consequently reduces the battery autonomy. On the other end, to operate the PA efficiently in terms of consumed power, its signal transfer characteristic becomes strongly nonlinear. So, it is evident that a compromise between power efficiency and

linearity of the PA must be achieved. This compromise is a major concern on the design and performance of an overall communication system.

With the evolution of the computation technologies and the advent of virtual communication system simulators, the need for PA models that could mimic the PA behaviour appeared. The PA models allow the characterization of their impact on system's performance. Two different approaches can be used to reach these models. One can start from the PA circuit design and build a virtual circuit from it through voltages or currents Kirchoff's laws. Alternatively, one can adopt a black box methodology which, regardless of the amplifier circuit, tries to build a mathematical function that approximates the PA response to a given set of inputs. The first approach, which might be a good solution for a circuit with a small number of active elements, rapidly becomes unpractical if their amount increases, since the number of equations leads to a mathematical problem hard to solve. So, the alternative is to find a PA model that does not have the same complexity of the equivalent circuit model while still being able to approximate the PA response. This is the principle of the black box modelling or behavioural modelling approach. This modelling is distinguished from the equivalent circuit modelling because no parallelism can be made among the circuit being modelled and the model topology.

Another application of the behavioural models is its use on linearization. To alleviate the compromise between power efficiency and linearity on the PA design, some linearization techniques have been proposed throughout the last years [1.1,1.2]. One of the most discussed today is based on digital baseband predistortion. In nowadays communication systems there is a digital processor that is responsible for the baseband digital treatment of the information: coding, interleaving, spreading etc. The basic principle of digital predistortion techniques is to substitute the ideal signal to transmit by one signal somehow designed to become the ideal signal after the PA distortion effects. To be able to generate this predistorted transmission signal it is required a good inverse model of the amplifier, which is an equivalent problem of getting a good model of the amplifier.

As was stated above, the use of power amplifiers in nowadays communication systems is a key issue, since they have to be efficient, in order to extend portable devices' battery life, and also linear to accomplish the tight spectral masks imposed by the standards to allow a practical use of the spectra. This would not be a problem if the maximum device power efficiency would not be reached in a very nonlinear regime.

In order to reach a good compromise between efficiency and linearity, behavioural modelling plays a very important role either to allow the implementation of fast and accurate system level simulations or to promote the design of base-band digital pre-distorters that increase the PA efficiency.

An ideal power amplifier should work as a constant gain factor applied to the input signal

$$y(t) = G \cdot x(t) \tag{1.1a}$$

or also as linear gain only on the input signal bandwidth zone (an active filter):

$$y(t) = \int_{-\infty}^{+\infty} g(\tau)x(t - \tau)d\tau \tag{1.1b}$$

where $x(t)$, $y(t)$ are the input and output signals; G and $g(\tau)$ are the amplifier gain and impulse response respectively.

The power required to operate this "ideal" PA is only the power that is effectively added to the signal.

In this scenario the PA is a linear system in which the superposition principle holds. That is, if the system outputs to $x_1(t)$ and to $x_2(t)$ are known, then the systems response to any linear combination of these two inputs is also known:

$$\mathcal{L}[\alpha_1 x_1(t) + \alpha_2 x_2(t)] = \alpha_1 \mathcal{L}[x_1(t)] + \alpha_2 \mathcal{L}[x_2(t)] \tag{1.2}$$

Linear systems have been extensively studied over the last decades, and their identification in the presence of nonlinear distortion continues to be an important investigation field [1.3-1.5]. However a simple linear based model is of no practical use when trying to estimate the distortion effect on the communication system, since a linear system does not account for distortion.

Several approaches have been used to model the nonlinear behaviour of a PA. Probably the simpler ones are the static representations in which the output signal $y(t)$ is obtained as:

$$y(t) = F[x(t)] \tag{1.3}$$

Usual representations of the nonlinear function F are polynomial series transfer functions or functions based on hyperbolic tangent like approximations. Being very simple in their nature these methods are only capable of representing the memoryless or static part of the PA nonlinearity phenomena.

In order to account for the PA dynamic effects lots of different approaches have been proposed in the last years. They can be grouped into two main categories, the artificial neural networks (ANN) and the polynomial approximators.

With the new computational capabilities ANNs are easy tools to work with. This is one of their major advantages. However, their coefficient extraction procedure – the learning process – relies on a nonlinear optimization procedure that tries to minimize a certain error function. After the convergence of this process the ANN presents good approximation results for the input signal class used in coefficient determination, but has no guaranteed modelling accuracy for signal classes different from the one used for extraction.

On the other hand, polynomial models and, in particular, Volterra series models are linear in their parameters and so they can be extracted in a more straightforward way. Expression (1.4), represents the digital formulation of the Volterra series.

$$\begin{aligned} y(s) = & \sum_{i_1=0}^{M} h_1(i_1)x(s-i_1) \\ & + \sum_{i_1=0}^{M}\sum_{i_2=0}^{i_1} h_2(i_1,i_2)x(s-i_1)x(s-i_2) + \cdots \\ & + \sum_{i_1=0}^{M}\sum_{i_2=0}^{i_1}\cdots\sum_{i_N}^{i_{N-1}} h_N(i_1,i_2,\ldots,i_N)x(s-i_1)x(s-i_2)\ldots x(s-i_N) \end{aligned} \tag{1.4}$$

where, M represents the number of time samples considered and N the nonlinear order. In this expression the $h_x(.)$ parameters linearity is shown since each of them multiplies a combination of input samples.

However, due to the overlapping of a large number of different Volterra series terms, as can be seen in (1.4), it is hard to determine exactly each single coefficient. So, the coefficients are usually determined with the use of some linear estimation technique like least square error minimization. This way, there are no guarantees that the coefficients

obtained are also optimum for a different kind of input signal. Or equivalently, the set of coefficients that model a given system might vary with different signals used for model extraction. If a method for separate identification of each coefficient is possible then the true coefficient value could be determined independently of all the others.

1.1 Objectives

The main goal of this thesis is to formulate a procedure for the orthogonal Volterra series parameter extraction. So, a rearrangement of the Volterra series that allows the separation of all its components (orthogonality) is searched. Note that this orthogonality is achieved only for a particular input signal statistics, since the input signal has impact on the output signal statistics. With this orthogonal model formulation, the coefficient determination is done in a one to one way and, finally, these orthogonal series coefficients are transformed into the conventional Volterra series parameters.

With this approach it is guaranteed that the best polynomial system approximator, up to a given order N and memory span M is obtained.

The goal of this thesis is to provide some contributions to the study of the distortion impact on a communication system.

The objectives defined for this thesis are:

- To formulate a behavioural model topology and the corresponding extraction procedure, with a well known mathematical background so that:
  - The coefficient extraction is unique and straightforward
  - The model predictive capabilities are guaranteed.
- To propose a metric to evaluate the signal quality degradation in a nonlinear system;

1.2 Thesis Description and Original Contributions

To reach the goals above presented this thesis is organized in the following way. Chapter 2 provides a state of the art description on power amplifier behavioural modelling. This chapter starts with a brief description of the main characteristics that must be modelled by the PA to then present a description (not exhaustive since there are a very

large number of different approaches) of the most important works on this area. The description presented on this part of the thesis follows closely a recent paper on this subject [1.6]. In the second part of this chapter some previous orthogonal approaches to Volterra series based behavioural modelling are presented.

Chapter 3, one of the most important in this thesis, presents the formal derivation of the orthogonal behavioural model proposed, the extraction procedure and correspondence between the orthogonal model's coefficients and the corresponding time domain Volterra series. This Chapter is followed by the practical validation results shown in Chapter 4, where the modelling procedure is tested in a wide range of different situations ranging from a simulated memoryless amplifier to a real PA with memory effects. These two chapters aggregate a set of original author contributions published in several national and international conference papers [1.7-1.12].

As an application of cross-correlation identification techniques, and output signal components identification, Chapter 5 presents the author contributions to evaluate the signal quality degradation in a nonlinear system due to noise and distortion. To come up with the definition of Noise and Distortion Figure, the Best Linear Approximator of a nonlinear system is computed. The Best Linear Approximator is derived using cross correlation identification techniques similar to the ones used to separate each component of the orthogonal model. This chapter ends with some application results of the concepts proposed. The value of this original work was recognized by one paper on a national conference [1.13], two papers published at the International Microwave Symposium [1.14,1.15], and an extended version of that work in the IEEE Transactions on Microwave Theory and Techniques [1.16].

To finish this thesis, Chapter 6 presents the conclusions of the work performed and also some guidelines for Future Work that can be carried on as a natural sequence to this thesis.

1.3 References

[1.1]. A. Zhu and T. J. Brazil, "An adaptive Volterra predistorter for the linearization of RF high power amplifiers," in *IEEE MTT-S Int. Microwave Symp. Digest*, June 2002, pp. 461-464.

[1.2]. M. O'Droma, E. Bertran, M. Gadringer, S. Donati, A. Zhu, P. L. Gilabert, and J. Portilla, "Developments in predistortion and feedforward adaptive power amplifier linearisers," *European Gallium Arsenide and Other Semiconductor Application Symposium Digest*, Oct. 2005,pp. 337-340.

[1.3]. Ai Hui Tan; K. R. Godfrey and H. A. Barker, "Design of computer-optimized pseudorandom maximum length signals for linear identification in the presence of nonlinear distortions," *IEEE Trans. on Instrumentation and Measurement*, Vol. 54, Issue 6, pp. 2513 – 2519, Dec. 2005.

[1.4]. R. Pintelon; E. Van Gheem, L. Pauwels, J. Schoukens and Y. Rolain,"Improved approximate identification of nonlinear systems," *Proceedings of 21st IEEE Conf. on Instrumentation and Measurement Technology*, May 2004, pp. 2183 – 2186.

[1.5]. R. Watanabe and K. Uchida, "A practical linear identification method based on geometric representation of time invariant continuous systems," *Proceedings of 42nd IEEE Conf. on Decision and Control*, 9-12 Dec. 2003 , pp.1126 – 1131.

[1.6] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. on Microwave Theory & Tech.*, vol. MTT-53, Issue 1, pp. 1150-1163, April 2005.

[1.7]. P. M. Lavrador, J. C. Pedro and N. B. Carvalho, "A new nonlinear quasi-orthogonal model extracted in the frequency domain," *5th National Conference on Telecommunications Digest,* Tomar, Portugal 2005.

[1.8]. P. M. Lavrador, J. C. Pedro and N.B. Carvalho, "A new nonlinear behavioral model extracted orthogonally," *XX Conf. On Design of Circuits and Integrated Systems Proc. CDROM*, Lisbon, Portugal, Nov. 2005.

[1.9]. P. M. Lavrador, J. C. Pedro and N. B. Carvalho, "A new Volterra series based orthogonal behavioral model for power amplifiers," *Asia & Pacific Microwave Conference Proceedings,* Suzhou, China 2005.

[1.10]. P. M. Lavrador, J. C. Pedro and N. B. Carvalho, "A new method for the orthogonal extraction of the Volterra series coefficients," *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits Proceedings CDROM,* Aveiro, Portugal, Jan. 2006, pp. 138-141.

[1.11] J. C. Pedro, P. M. Lavrador and N. B. Carvalho, "A formal procedure for microwave power amplifier behavioural modelling", in *IEEE MTT-S Int. Microwave Symp. Digest*, San Francisco, USA, 2006, pp. 848-851.

[1.12]. C. Siviero, P. M. Lavrador and J. C. Pedro, "A frequency domain extraction procedure of lowpass equivalent behavioural models of microwave PAs," *1st Integrated Microwave Integrated Circuits Conference Proceedings*, Manchester, U.K., Sept. 2006, pp. 253-256.

[1.13]. P. M. Lavrador, J. C. Pedro and N. B. Carvalho, "Underlying linear system identification of memoryless and dynamic nonlinear systems," *4th National Conference on Telecommunications Digest,* Aveiro, Portugal 2003.

[1.14]. P. M. Lavrador, N. B. Carvalho and J. C. Pedro, "Noise and Distortion Figure – an extension of noise figure definition for nonlinear devices," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, U.S.A, 2003, pp.2137-2140.

[1.15]. J. C. Pedro, N. B. Carvalho and P.M. Lavrador, "Modeling nonlinear behavior of bandpass memoryless and dynamic systems," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, U.S.A, 2003, pp. 2133-2136.

[1.16]. P.M. Lavrador, N. B. Carvalho and J. C. Pedro, "Evaluation of signal to noise and distortion degradation in nonlinear systems," *IEEE Trans. on Microwave Theory and Techniques,* - Vol. 52, Issue 3, pp. 813-822, March 2004.

2. State of the Art Description

2.1 Introduction

In this chapter an overview of the most relevant works that have been done (or are currently being done, at the time of this thesis) in the area of power amplifier behavioural modelling is presented. The goal of this overview presentation is to introduce the main issues on behavioural modelling and also to justify the modelling approach that has been followed on the work presented in this thesis.

This description will be divided into two main parts: the first one is devoted to present the state of the art in behavioural modelling in general; while the second one will be more focused on the Volterra Series (VS) describing some previous approaches to VS behavioural modelling especially some works on the optimal VS coefficients extraction.

The first section of this chapter starts by trying to identify the driving forces to behavioural modelling and the compromises that are required among them. Then a description of the PA phenomena that must be modelled is made. To conclude the first section of this chapter some examples of previous behavioural model approaches are

presented to illustrate different modelling capabilities. This part of the chapter follows closely the behavioural model comparison of [2.1].

The second part of the chapter, which is devoted to VS modelling, starts with a succinct historical introduction to the VS, to then describe different techniques that had been proposed to extract its coefficients. It is precisely on the coefficient extraction difficulties that is placed the emphasis of this section that ends with the description of prior orthogonal approaches to determine them.

2.2 Behavioural Modelling Overview

Along the chain from active transistor fabrication to the Power Amplifier circuit design and system level performance evaluation there are different requirements for modelling and characterization of the devices and/or systems. As usual in engineering problems, all those requirements involve a trade off between effort and performance.

For instance, from the measurement engineer point of view, an accurate knowledge of the device and also an exhaustive exploration of its different operation regimes would be desirable. However that exhaustive testing can imply a large number of measurements to be done, with probably different setups which may take too much time and/or cost to be performed.

The circuit designer requires a model that accurately describes the device, but that also allows fast Computer Aided Design (CAD) simulation in order to optimise the time required for simulations of different circuit configurations. On the other end of this compromise, there is the complexity of the model, the number of physical effects that are handled and, more important, the reliability of the model.

At the top of hierarchy, the system designer would like to verify the overall system's performance and to find the best trade off between linearity and power efficiency; modulation with varying envelopes (to optimize the transmission rates) and amplifiers with low frequency memory effects.

On each different stage a model is required to allow the design and simulation for performance evaluation of the proposed solution. However, since in each of the stages the designer is facing different problems, the model characteristics should also be different. Actually to perform a complete system simulation using the interconnection of the physical

model of each of its components is not a good solution due to the complexity/number of variables required and also to the large computation time required [2.2].

It is usual, especially on system level simulations, to rely on simplified models that are able to describe the input/output relation of a given (sub-)system without having to include all the physical information about its components. In some cases, for instance Travelling Wave Tube (TWT) amplifiers, there is effectively no relation between the device and an equivalent circuit. Thus, an empirical model based on the observation of input/output characteristics is developed to allow the simulation of that kind of devices.

Given the main driving forces for behavioural model use, it should be of no surprise that this has been a hot topic in the last years. However, the increasing interest on this subject was not accompanied by a solid theoretical work. Effectively, perhaps due to the empirical nature of the models, many have been proposed without references to similar on-going works. A recent paper [2.1] presented an overview of the main activities that have been done in this area. In that paper the formal background of system identification theory is discussed, and then different model approaches are classified according to their approximation capabilities.

A behavioural model is presented as an operator that intent to approximate the response of one system – which is a function, or a vector of functions – to a certain input excitation – again one function or vector of functions. In order to introduce some mathematical formalism it can be written, for a general case as:

$$y(t) = F\left[x(t), \frac{dx(t)}{dt}, \ldots, \frac{d^n x(t)}{dt^n}, \frac{dy(t)}{dt}, \ldots, \frac{d^m y(t)}{dt^m}\right] \tag{2.1}$$

This equation explicits that the output $y(t)$ is a function of the input $x(t)$ and of its derivatives up to the n^{th} order, and of the output derivatives up to the m^{th} order.

Since behavioural models are intended to be used in digital computers, (2.1) can be re-written using its discrete time equivalent:

$$y(s) = F[x(s), x(s-1), \ldots, x(s-n), y(s-1), \ldots, y(s-m)] \tag{2.2}$$

In which it is stated that the output at the time instant s (that effectively corresponds to sT_s), is a function of the present input sample $x(s)$, of its n past samples, and also a

function of the output *m* past samples. Equation (2.2) can be recognized as a nonlinear infinite impulse response (*NIIR*) filter [2.3].

An interesting result from system identification theory states that any continuous, stable, causal and of finite memory system – which is a good framework for a general PA – can also be represented with any small error using a non recursive structure. This means that a different functional form can be found:

$$y(s) = F_{NR}[x(s), x(s-1),..., x(s-M)] \tag{2.3}$$

M represents the number of past input samples actually required to represent the current output sample and is called the memory depth, or memory span of the system. The function $F_{NR}[.]$ can be identified as a nonlinear finite impulse response (NFIR) filter. The non-recursive approximation of any system may have some disadvantages like the possible large number of coefficients to achieve a negligible error. However, it has one guaranteed main advantage: it is always unconditionally stable. This is why most of the modelling approaches adopt this type of formulation. Additionally, it is easier to determine the coefficients of a feed-forward structure than the ones of a recursive structure.

There are several ways to implement the function $F_{NR}[.]$. The two more common are the multidimensional polynomial and the time delay artificial neural networks (ANN). The reasons for the use of either of these two solutions are their formal mathematical support and also the fact that they lead to a straightforward implementation. In the case of a multidimensional polynomial it replaces $F_{NR}[.]$ and so (2.3) takes the form of:

$$\begin{aligned} y(s) = & \sum_{i=0}^{M} P_{1,i} x(s-i) \\ & + \sum_{i_1=0}^{M} \sum_{i_2=0}^{M} P_{2,i_1 i_2} x(s-i_1) x(s-i_2) \\ & + \sum_{i_1=0}^{M} \sum_{i_2=0}^{M} \sum_{i_3=0}^{M} P_{3,i_1 i_2 i_3} x(s-i_1) x(s-i_2) x(s-i_3) + \cdots \end{aligned} \tag{2.4}$$

in which it is shown the dependence of $y(s)$ on a series of multi-linear terms. Actually it can be seen that, regardless of the order of each term, there is a linear relation between the output and each of the n^{th} order product terms. Also, a linear relation can be found for the

coefficients P_n. These linear properties made easy the formal study of these multidimensional polynomials in terms of their convergence and approximation capabilities..

There are two well known particular cases of this approach. (i) If $F_{NR}[.]$ is approximated in a Taylor series sense, then the multidimensional polynomial is known as the Volterra series and has the property of producing the optimal approximation (in a uniform error sense) for an expansion near the point where it was extracted. It is thus especially good for approximation in the small-signal (or mildly nonlinear) regime and quite bad for large signal regimes. (ii) If $F_{NR}[.]$ is approximated by an Hermite polynomial, then (2.4) is known as the Wiener series, and produces approximation results optimal (in a mean square error sense) in the vicinity of the extraction input power level and for that kind of input signals. Thus it is amenable for modelling stronger nonlinear systems, when the input signal characteristics can be considered close to the ones of the white noise.

In the case where $F_{NR}[.]$ is approximated by an ANN (2.3) takes the form of:

$$\begin{aligned} u_k(s) &= \sum_{i=0}^{M} [w_k(i)x(s-i)] + b_k \\ y(s) &= b_0 + \sum_{k=1}^{K} w_o(k) f[u_k(s)] \end{aligned} \tag{2.5}$$

where $w_k(i)$, $w_o(k)$, b_0 and b_k are weighting coefficients and bias parameters to be determined during the extraction procedure (ANN training), while $f[.]$ is a predetermined function, known as activation function. The formulation shown in (2.5) is a simplified version of a class of ANNs known as the multi-layer perceptron with a single hidden layer. In [2.4,2.5] it was shown that this type of ANN has universal approximation capabilities.

It can be easily shown that if the ANN's activation function is bounded then the approximation results of the ANN are also bounded. This characteristic is a great advantage when compared to the known divergence problems of the polynomials. On the other hand, the main disadvantage of the ANN is that no direct relation can be made between the output and its coefficients, thus one has to rely solely in an optimization process to determine them. This implies that one has no guarantee that the optimal solution is found. This poses a bigger problem since nothing can be said on what will be the impact

(in terms of approximation quality) of increasing or decreasing the number of model parameters.

At this point it is worth stressing out that there are actually two big families of models: the "bandpass models" that process RF-modulated carriers and the "lowpass equivalent models" that process only the envelope information regardless of the carrier frequency. Recalling that a PA is a device intended to amplify a signal like:

$$x(t) = \mathrm{Re}\left[r(t)e^{j[\omega_c t + \phi_r(t)]}\right] \tag{2.6}$$

that consists of a carrier modulated by an information signal $[r(t), \phi_r(t)]$, then a model able to process $x(t)$ – the bandpass model – or a model that processes the envelope of the carrier $\tilde{x}(t)$ – a low pass equivalent model – can be conceived.

This difference is actually quite important due to the large disparity between the time scales of the carrier and the envelope (e.g. in WCDMA the envelope bandwidth is near 5MHz while the carrier is near 2GHz). So a model intended to process simultaneously the carrier and the envelope can be hard to implement in practice due to the very large number of time samples that will be required to accommodate a reasonable number of envelope signal periods sampled at a sufficient rate to represent the carrier.

2.2.1 Main Effects to be Modelled by a PA Behavioural Model

In a recent publication [2.6] the dynamic effects of PAs were divided into three categories according to their ability of representing PA memory effects: the static or memoryless, the PAs with linear memory and the PAs with nonlinear memory.

The first type – static PAs – can be represented as shown in Figure 2.1.

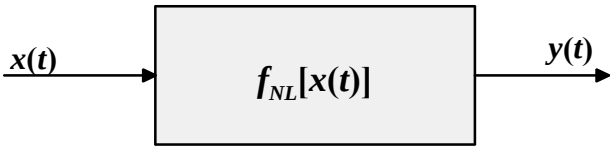


Figure 2.1 – Schematic representation of a memoryless amplifier.

In this figure it is shown that the output $y(t)$ is obtained as a function of the present input $x(t)$ regardless of its past (derivatives in the case of continuous signals). If a Taylor series (or other polynomial form) expansion of $f_{NL}[x(t)]$ is considered:

$$f_{NL}[x(t)] = \sum_{n=0}^{\infty} \alpha_n x(t)^n \tag{2.7}$$

then the first and third order frequency domain Volterra series kernels of this system can be written as [2.6,2.7]:

$$\begin{aligned} H_1(\omega) &= \alpha_1 \\ H_3(\omega_1, \omega_2, \omega_3) &= \alpha_3 \end{aligned} \tag{2.8}$$

It can be shown that this is a reasonable approximation of a PA processing a very narrow input signal, for which its transfer characteristics remain essentially constant inside the signal bandwidth.

However, if it is considered an input signal bandwidth that is not negligible compared to the amplifier pass band, then the effects of the input and output matching networks must be taken in account. In this case a cascade of the memoryless nonlinearity between to dynamic linear filters is obtained.

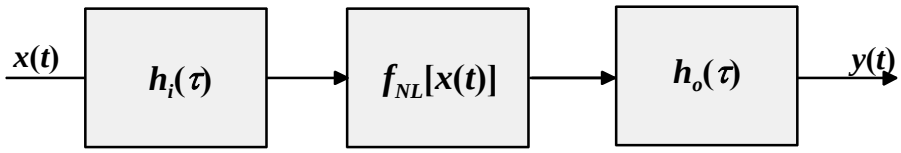


Figure 2.2 – Schematic representation of nonlinear amplifier with linear memory.

In Figure 2.2 the input and output filters are represented by their impulse responses $h_i(\tau)$ and $h_o(\tau)$, respectively. For this case and considering the same Taylor series expansion of (2.7) the equivalent Volterra series description of this system is [2.6,2.7]:

$$\begin{aligned} H_1(\omega) &= H_I(\omega) \cdot \alpha_1 \cdot H_O(\omega) \\ H_3(\omega_1, \omega_2, \omega_3) &= \alpha_3 \cdot H_I(\omega_1) \cdot H_I(\omega_2) \cdot H_I(\omega_3) \cdot H_O(\omega_1 + \omega_2 + \omega_3) \end{aligned} \tag{2.9}$$

where $H_I(.)$ and $H_O(.)$ are the frequency domain representations of the filters $h_i(\tau)$ and $h_o(\tau)$, respectively.

In some cases it can be observed a different type of dynamics in PAs: nonlinear memory effects. Those are memory effects that are only visible when the amplifier enters it's nonlinear regime of operation. This type of effects can be modelled by a different topology as shown in Figure 2.3.

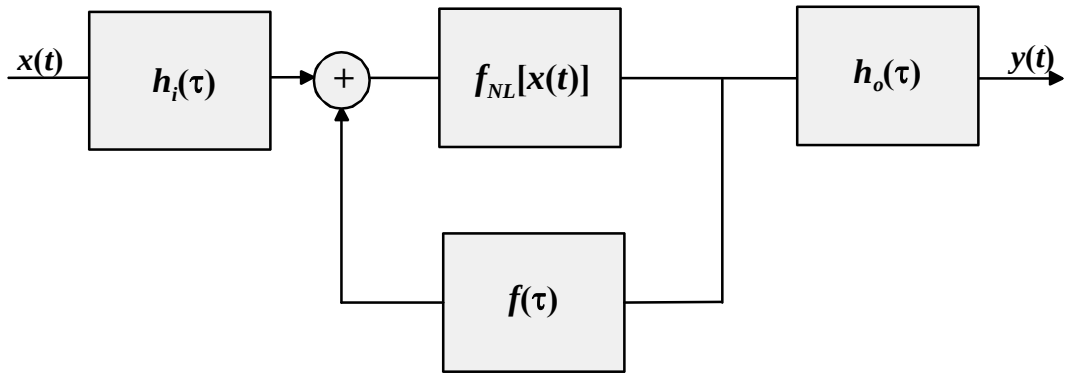


Figure 2.3 – Schematic representation of a nonlinear amplifier with nonlinear memory proposed by Pedro *et al.*

Considering once again the Taylor series expansion of (2.7) for $f_{NL}[x(t)]$, the first and third order Volterra kernels of this system are [2.6,2.7]:

$$H_1(\omega) = I(\omega) \cdot \frac{\alpha_1}{D(\omega)} \cdot O(\omega) \tag{2.10a}$$

$$\begin{aligned} H_3(\omega_1, \omega_2, \omega_3) = {} & \frac{I(\omega_1)I(\omega_2)I(\omega_3)}{D(\omega_1)D(\omega_2)D(\omega_3)} \frac{O(\omega_1 + \omega_2 + \omega_3)}{D(\omega_1 + \omega_2 + \omega_3)} \cdot \\ & \left\{ \alpha_3 + \frac{2}{3}\alpha_2{}^2 \left[\frac{F(\omega_1 + \omega_2)}{D(\omega_1 + \omega_2)} + \frac{F(\omega_1 + \omega_3)}{D(\omega_1 + \omega_3)} + \frac{F(\omega_2 + \omega_3)}{D(\omega_2 + \omega_3)} \right] \right\} \end{aligned} \tag{2.10b}$$

where $D(\omega) = 1 - \alpha_1 F(\omega)$.

Contrary to the Volterra descriptions of the last examples, in this equation the nonlinear dynamic effects are visible in the product of $\alpha_2{}^2$ by a rational of the feedback filter transfer function.

An alternative model topology was also proposed by Vuolevi *et al*. [2.8,2.9] that is also capable of modelling nonlinear dynamic effects.

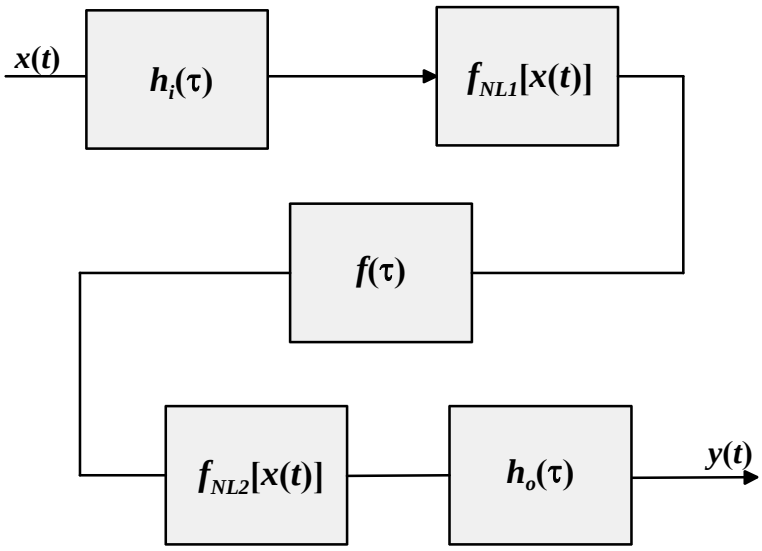


Figure 2.4 – Schematic representation of a nonlinear amplifier with nonlinear memory. proposed by Vuolevi *et al.*

In this model topology the nonlinear dynamic effects are modelled by the nonlinear mixtures of the dynamic response of the filter $f(\tau)$ that are created in $f_{NL2}[]$. Note that since the filter $f(\tau)$ is placed after $f_{NL1}[]$ it processes baseband components created by the first nonlinearity that are then up converted by $f_{NL2}[]$.

If we assume that $f_{NL1}[]$ is equal to $f_{NL2}[]$, then this model topology is a simplified version of the one of Figure 2.3.

2.2.2 Different Behavioural Model Capabilities

In this section some illustrative examples of the more common behavioural model approaches are described. Different model examples are presented according to their capabilities of modelling the different phenomena presented in the last section.

Note that it is not the aim of this text to present an exhaustive description of all the behavioural modelling approaches. Given that they are so many, its description would go far beyond the scope of this thesis.

Memoryless Models

The models with the simpler structure are also the ones with less dynamic predictive capabilities – the memoryless models. The most commonly cited model of this kind is the complex polynomial series:

$$\tilde{y}(t) = \sum_{n=0}^{N-1} a_{2n+1}\tilde{x}(t)|\tilde{x}(t)|^{2n} \tag{2.11}$$

where $\tilde{x}(t)$ is the complex envelope input and a_{2n+1} are the complex coefficients used. It is seen in (2.11) that in this approach the output envelope $\tilde{y}(t)$ is dependent only on the present value of the input envelope, thus it is impossible to represent any memory effect with a model formulation like this one.

Another well known memoryless model is the Saleh model, [2.10] in which:

$$r_y[r_x(t)] = \frac{\alpha_r r_x(t)}{1 + \beta_r [r_x(t)]^2} \tag{2.12}$$

and

$$\phi_y[r_x(t)] = \frac{\alpha_\phi r_x(t)}{1 + \beta_\phi [r_x(t)]^2} \tag{2.13}$$

where r_x, ϕ_x, r_y and ϕ_y are the amplitude and phase parts of the input and output complex envelope signals, respectively. α_r, β_r, α_ϕ and β_ϕ are parameters used to fit the modelled and measured transfer curves, from Amplitude Modulation to Amplitude Modulation (AM-AM) and from Amplitude Modulation to Phase Modulation (AM-PM). Despite this modelling approach intends to model amplitude and phase, both of them are once again only dependent on the instantaneous input complex envelope amplitude, and thus this model is considered a memoryless model.

Models of this type – memoryless – are well suited for modelling systems in which the input and output filters' ($h_i(\tau)$ and $h_o(\tau)$ in Figure 2.3) bandwidth is considerably wider than the bandwidth of the signal processed by the system, so that those filters can be considered flat. Additionally, the feedback path filter $f(\tau)$ should also be flat so that the nonlinear mixing through the feedback path doesn't create memory effects in the system.

Models addressing linear memory

When the signal's bandwidth approaches the bandwidth of the system, then its response can no longer be considered flat – it presents linear memory effects. This way the memoryless models (extracted with a single CW excitation) are no longer valid approximators for these systems. Since the approximation problem appeared from the bandwidth increase, one of the proposed solutions was to consider AM-AM and AM-PM curves dependent on the input frequency. This was the approach of Saleh frequency dependent model [2.10] intended to be used on TWT amplifiers. Saleh proposed a new model formulation in which the complex envelope output amplitude and phase are dependent on the input complex envelope amplitude as previously, but also on the carrier frequency. This dependence of the AM-AM and AM-PM curves on the carrier frequency can be understood as the modelling of a linear filter connected in series with the previous frequency independent nonlinearity. To understand the impact of the input and output filters, let's focus on the schemes of Figure 2.1 and Figure 2.2. In the memoryless case the output complex envelope can be written as:

$$\tilde{y}(t) = f_{NL}[\tilde{x}(t)] \tag{2.14}$$

If the effect of the input filter is now considered, in a configuration usually known as Wiener two box model, the complex output will be given by:

$$\tilde{y}(t) = f_{NL}[h_i(t) * \tilde{x}(t)] \tag{2.15}$$

The input filter can change the amplitude of the signal entering the nonlinear block, feeding the nonlinearity with different input power levels. Thus the AM-AM and AM-PM curves can be shifted horizontally (along the input power axis) by changing the coefficients of the input filter.

On the other hand if the effect of the output filter is considered separately, the two-box Hammerstein model is obtained. In this case the output complex envelope will be given by:

$$\tilde{y}(t) = f_{NL}[\tilde{x}(t)] * h_o(t) \tag{2.16}$$

From this expression it can be seen that the output filter impulse response $h_o(\tau)$ is convolved with the nonlinearity output and thus can control the output amplitude. Thus the output filter can shift the AM-AM and AM-PM curves vertically (along with the output power/output phase axis).

The reasoning above presented states that both configurations are essentially different, and thus cannot be interchanged. Actually, it is quite frequent the usage of both configurations simultaneously in a three box configuration: filter – memoryless nonlinearity – filter. This structure is also known as the Wiener-Hammerstein model.

One of the first published works using this topology was one work of Blachman in the 60's [2.11]. In these earlier studies a framework was created for the study of nonlinear bandpass systems. Blachman identified each of the nonlinearity output autocorrelation function components: the signal X signal, the noise X noise and the signal X noise [2.12]. With these definitions, ways of identifying and separating the output signal from the output noise components were developed in order to characterize the signal quality degradation in the nonlinearity. Furthermore, studies of the interference impact of two signals of different amplitudes processed in the same PA were conducted [2.13].

Further examples of three-box models are the Poza-Sarkozy-Berger [2.14] and the Abuelma'atti models [2.15].

Different modelling approaches that also lead to models with these predicting capabilities are the ones based on polyspectral higher order statistics proposed by C. Silva *et al.* in [2.16]. In these works it is assumed that the nonlinear FIR filters can be represented by one-dimensional systems – which forces the systems to be represented by a two box model (either Wiener or Hammerstein configuration), in which the nonlinearity is the measured AM-AM and AM-PM curve. The main advantage of this approach is that it allows the model extraction with one CW test (to determine the AM-AM and AM-PM static curves) followed by an envelope test (to determine the cascaded filter). However this configuration is incapable of representing the systems' nonlinear memory, as will be seen in the next sub-section. Other example of a three box model is the one proposed by Ibnkahla et al [2.17], that uses an ANN to approximate the three box structure.

Models Addressing Nonlinear Memory

Despite all the model formulations presented above, there are some types of effects that can not be modelled by any of the models already presented. For instance, an amplifier in which the output $\tilde{y}(t)$ depends, not only on the present value of the input complex envelope $\tilde{x}(t)$, but also on its past values requires a different model structure.

Back in 2000, Fang *et al*. [2.18] proposed a recursive neural network model for a 1 GHz bandpass amplifier. Unfortunately the ANN was trained and tested with the same type of unmodulated data. This way the ANN training process can not capture any envelope dynamic effects since they were not excited by the training data. In this case the adopted extraction procedure (basically the excitation choice) does not allow taking full profit from the general predictive capabilities of the adopted ANN structure.

A different approach has been proposed by Mirri *et al*. [2.19] and further developed by Ngoya *et al*. [2.20] and Soury *et al*. [2.21]. This strategy consists in an application of the nonlinear integral model (NIM) of Filicori *et al*. [2.22] and is based on the assumption that the signal can be nonlinearly processed in the system in a static way while the dynamic effects introduced can be described in a linear way. In this strategy the systems' response is expanded in a Taylor series around a predetermined nonlinear operating input $x_0(t)$ in the following way:

$$y(t) \approx \sum_{q=1}^{Q} f_q\left[x_0(t), \tau_q\right] \cdot x\left(t - \tau_q\right) \tag{2.17}$$

this formulation can lead to the definition of a nonlinear impulse response, dependent on the static approximation point $h[x_0(t),\tau]$

$$y(t) \approx \int_0^{\infty} h[x_0(t), \tau] \cdot x(t - \tau) d\tau \,. \tag{2.18}$$

A different approach that also leads to a parallel Hammerstein topology was proposed by Ku [2.23]. Ku intended to model the memory effects shown by power amplifiers when excited by two tone RF signals. In that work the authors started with a usual AM-AM/AM-PM memoryless polynomial representation of the nonlinearity (similar to the one in (2.11)) and imposed coefficient dependency with frequency, obtaining, this way, a model like:

$$\tilde{y}(t) = \sum_{n=0}^{N-1} a_{2n+1}(\omega) \cdot \tilde{x}(t) \cdot |\tilde{x}(t)|^{2n} \tag{2.19}$$

where ω_m represents the frequency of the envelope sinusoid, which is the envelope equivalent of the particular two tone input.

A heuristic parametric approach has been proposed by Asbeck *et al*. [2.24] and continued by Draxler *et al*. [2.25]. The main idea of this approach was the extension of the AM-AM/AM-PM modulation by assuming that for an amplifier with long term memory effects its transfer characteristics will no longer be solely dependent on the instantaneous input envelope amplitude but also on a parameter that is a function of the envelope past inputs. If the PA transfer characteristic is represented by *G* which used to be considered as a function of $r_x(t)$: $G[r_x(t)]$; this work considers *G* as a function of the present input envelope and also of a parameter $\tilde{z}(t)$ dependent on the past input envelope samples: $G[\tilde{z}(t), r_x(t)]$. Then the authors expand $G[\tilde{z}(t), r_x(t)]$ by a first order Taylor series. This way the PA is modelled by a complex gain that is a nonlinear function of the instantaneous envelope amplitude and also of a parameter obtained by linear filtering of the input envelope past samples.

In a recent publication Dooley *et al*. [2.26] proposed a recursive model based on a IIR filter with a new approach for improving the coefficient determination of the structure. Starting with a vector of input and output data it was possible to develop an algorithm that uses the measured output sample at each process' iteration for coefficient determination (and not the one computed with the estimated coefficients). This way the convergence and accuracy of the coefficient extraction are increased. The results obtained with this modelling strategy are compared with a conventional 3rd order VS model and a model approximation improvement was shown.

A more formal approach has been followed by Brazil and his group. In 2003 they have proposed a least square error extraction procedure for a hybrid time and frequency domain Volterra representation of a RF bandpass PA [2.27]. Despite nothing is said on the order of the model extracted nor on the number of time delays considered, the fact is that remarkably good results have been achieved with a model able to reproduce the variations in the third and even the fifth order Inter-Modulation Distortion (IMD). Recently, Zhu *et al*

[2.28], from the same research group, suggested a low pass equivalent model based on the Volterra Series written on discrete time domain. The main advantage of this work is that it is based on a well known framework – the Volterra Series – despite the limitations of applicability to mildly nonlinear PA's. This model is no-longer a one-dimensional arrangement, but a true multidimensional approximation of the system. This work had the merit of proposing an unconventional arrangement for the Volterra series – the V-vector algebra which simplified the model implementation. Since this work was devoted to address nonlinearity compensation techniques (pre-distortion), the authors proposed an adaptive learning process for the coefficient estimation instead of considering a more straightforward coefficient extraction. These works have been continued and recently the author has presented new methods to reduce the number of coefficients of the Volterra series [2.29,2.30].

This work, well grounded on the Volterra series field constitutes a general lowpass equivalent approach. However, the suggested adaptive coefficient determination and the overlapping existent between different model terms are still issues that can be improved in order to reach the best Volterra series based low pass equivalent model.

Actually, in order to optimally determine the Volterra series coefficients, some strategies should be adopted to enforce the separability among all the model terms and to obtain a well conditioned set of equations for coefficient determination. In the next section some previous works on efficient determination of Volterra series coefficients will be described.

2.3 Volterra Series Modelling

The Volterra series is so called in recognition of the work of the mathematician Vito Volterra. Around the year 1880, Volterra studied a series, as a generalization of the Taylor series expansion of a function, and later he lectured its application to the study of certain differential and integro-differential equations in 1912. Fréchet has used Volterra's results and added some more rigorous mathematical foundation. Specifically Fréchet proved that "any continuous functional can be represented by a series of functionals of integer order [equivalent to Volterra functionals] whose convergence is uniform in all compact sets of continuous functions". This result is a generalization of the Weierstrass theorem which

states that every continuous function can be represented by a series of polynomials that converges uniformly in every compact space of points [2.3]. These results were later used by Wiener who first used them to the study of nonlinear systems [2.31].

The Volterra series is a mathematical approximation of a general time invariant, continuous system written as

$$y(t) = \int_{-\infty}^{+\infty} h_1(\tau_1)x(t-\tau_1)d\tau_1 + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \cdots$$
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} h_n(\tau_1,\tau_2,\ldots,\tau_n)x(t-\tau_1)x(t-\tau_2)\ldots x(t-\tau_n)d\tau_1 d\tau_2 \ldots d\tau_n \tag{2.20}$$

in which for $n = 1, 2, \ldots$

$$h_n(\tau_1,\ldots,\tau_n) = 0, \qquad \textit{for any} \qquad \tau_j < 0. \tag{2.21}$$

With the growing use of computational calculation and simulation tools the discrete form of the Volterra series is of major importance:

$$\begin{aligned} y(s) = & \sum_{m_1=0}^{M} h_{m_1} x(s-m_1) \\ & + \sum_{m_1=0}^{M}\sum_{m_2=0}^{m_1} h_{m_1,m_2} x(s-m_1)x(s-m_2) + \cdots \\ & + \sum_{m_1=0}^{M}\sum_{m_2=0}^{m_2}\cdots\sum_{m_N}^{m_{N-1}} h_{m_1,m_2,\ldots,m_N} x(s-m_1)x(s-m_2)\ldots x(s-m_N) \end{aligned} \tag{2.22}$$

where M represents the length of the memory span, that is the number of past samples that can contribute to the output at the time sample s; and N the nonlinear order of the model considered. This equation is written in a triangular form according to the notation adopted in [2.32].

In addition to the solid mathematical groundwork, the main advantage of a Volterra series type model is that the kernel value determination is a linear problem. That is easily shown [2.32], in the case of (2.22), by considering a record of K input samples $x(0)$, $x(1)$,

…, $x(K)$ and output samples $y(0)$, $y(1)$, …, $y(K)$ and straightforwardly write a linear matrix equation where the unknowns are the kernel values.

$$Y = HX \tag{2.23}$$

where

$$Y = [y(0)\ \ y(1)\ \ \dots\ \ y(K)] \tag{2.24}$$

$$H = [h_0\ \ h_1\ \ \dots\ h_M\ \ h_{0,0}\ \ h_{1,0}\ \ h_{1,1}\ \ \dots\ h_{M,\dots,M}] \tag{2.25}$$

and

$$X = \begin{bmatrix} x(0) & x(1) & x(2) & \dots \\ 0 & x(0) & x(1) & \dots \\ 0 & 0 & x(0) & \dots \\ \vdots & \vdots & 0 & \\ x^2(0) & x^2(1) & & \\ 0 & x(0)x(1) & & \\ 0 & x^2(0) & & \\ 0 & 0 & & \\ \vdots & \vdots & & \\ x^N(0) & & & \\ 0 & & & \\ \vdots & & & \end{bmatrix} \tag{2.26}$$

It is now clear that if the number of recorded samples K is such that **X** is a square matrix and invertible then the kernel values **H** are obtained by:

$$H = YX^{-1} \tag{2.27}$$

However if K doesn't match the required size to make **X** square (either because it is larger or smaller), or if **X** is not invertible, then one has to rely on some least mean squares technique to obtain approximations to the kernel values. The problem of matrix **X** inversion can, in some situations, be solved with the use of the pseudo-inversion.

The proposed method for kernel determination states that it is possible to obtain the kernel values in a linear way, provided that the input is chosen with a certain length K and with certain properties to ensure that $\mathbf{X}$ is invertible. However, please note that the kernel determination using this direct formulation leads to a problem with very high dimensions for any case of practical interest, since the number of coefficients of an n^{th} degree kernel with memory span M is of the order of $(M+1)^n$. This can lead, and usually does, to difficulties in the solution of (2.27). Some published results on Volterra series behavioural modelling [2.33] adopt models of order 3 and memory span 1 to keep a low number of coefficients to be determined, while [2.28] uses a fifth order Volterra model without specifying the time delay considered.

A solution to circumvent the high number of coefficients described above is to introduce an approximation based on the expansion of the kernels using known functions. Describing the triangular form kernels using a series expansion will (in principle) decrease the number of coefficients required for its representation [2.32]. This text will not describe with further detail this approach, since it is out the scope of this thesis. However, it is an area that still raises interest in the scientific community as some recently published papers confirm [2.34,2.35].

Other identification techniques have also been proposed to determine general Volterra kernels. One example is the identification using input pulses, which is not very useful for RF power amplifier modelling due to the difficulties to create and measure pulses narrow enough to perform the kernel extraction. One technique that has been quite studied is the kernel identification using Gaussian white noise excitation [2.32,2.36]. This technique can be understood as an extension of the cross-correlation technique for the identification of linear systems, in which the input response can be computed as the cross correlation between the input and output of the system when the input is real, stationary Gaussian white noise with zero mean and power A:

$$\begin{aligned} h(\tau) &= \frac{1}{A} \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} y(t)x(t-\tau)dt \\ &= \frac{1}{A} E[y(t)x(t-\tau)] \end{aligned} \tag{2.28}$$

A similar analysis can be performed to determine the nonlinear impulse response of order n, exemplified by the second degree kernel:

$$h_2(\tau_1, \tau_2) = \frac{1}{2A^2} E[y(t)x(t-\tau_1)x(t-\tau_2)] \quad , for\ \tau_1 \neq \tau_2 \tag{2.29}$$

This equation however is only valid for $\tau_1 \neq \tau_2$, since for white Gaussian variables $E[x^2(t)]$ is non null, preventing this way the simplifications made to reach (2.29). Usually this problem is circumvented by either of two ways: (i) not using true white noise or (ii) obtaining the values of $h_2(\tau,\tau)$ by continuous extension of the values of $h_2(\tau_1,\tau_2)$ [2.32]. Either of these approaches can be valid in given circumstances however in chapter 3 it will be shown that the determination of diagonal values of h_n can cause difficulties and is of major importance. Another issue of using this method is the problem of separating different order terms as, for example, $h_1(\tau)$ and $h_3(x,x,\tau)$ that appear together in a nonlinear (3rd or higher) system case [2.32].

Another common approach for kernel determination is based on frequency response (FR) measurements. Again the underlying idea of nonlinear FR identification is based on the simpler concept used for linear system identification: if a stable, causal and linear system is considered, it's transfer function $H(j\omega)$ can be determined by measuring the magnitude and phase of the steady state response to the input $u(t) = \cos(\omega t)$. Actually two values of $H(j\omega)$ are determined since $H(-j\omega)$ is computed as the complex conjugate of $H(j\omega)$.

The determination of the second order kernel using one tone signal as excitation is incomplete since $H(j\omega_1, j\omega_2)$ can not be determined. Thus a two tone input signal is required for the determination of the second order kernel. This analysis is expandable to higher order kernels and thus a n^{th} tone input excitation is required to measure the n^{th} order kernel.

This strategy for kernel determination also leads to some problems [2.31,2.32,2.37] since there are different order terms that contribute to the same frequency component, needed to determine both of them. This can be understood considering the polynomial system case where the higher degree homogeneous parts of the system, also contribute to steady state response terms at the same frequencies as the lower degree terms (e.g. 3rd order correlated components). As a simple example consider the one tone input $u(t) = 2A\cos(\omega t)$ and a system composed only of degree 1 and 3 terms. The steady state output can be written as:

$$\begin{aligned} y_{ss}(t) &= 2A|H_1(j\omega)|\cos(\omega t + \angle H_1(j\omega)) \\ &+ 6A^3|H_3(j\omega, j\omega, -j\omega)|\cos(\omega t + \angle H_3(j\omega, j\omega, -j\omega)) \\ &+ 2A^3|H_3(j\omega, j\omega, j\omega)|\cos(3\omega t + \angle H_3(j\omega, j\omega, j\omega)) \end{aligned} \tag{2.30}$$

it is clear that there are two terms at frequency ω, and that the degree-3 homogeneous part of the system contributes to the frequency components needed to determine $H(j\omega)$.

Different homogeneous degree terms contributing to the same frequency bring up problems for determining the symmetric transfer functions from its evaluations. In order to deal with these problems some assumptions are required. However, they can depend on the type of system being studied [2.32]. One advanced solution is the use of cyclostationary inputs that simply eliminate this problem [2.38]. This solution will be discussed later when orthogonal extraction methods are described.

Schouckens and Pintelon [2.39] have been particularly active on the area of Frequency Response Functions (FRF) determination. In a 2001 publication of this group Vanhoenacker *et al*. [2.40] advanced a method for properly designing multisine excitations that allow detecting and qualifying the nonlinear distortions while measuring the FRF. This technique consisted in choosing a multisine in which not all the harmonics are excited. The non excited frequencies are used to detect and quantify the nonlinear effects.

However the special multisine design for Volterra Series identification was started long before. Starting in the early 80's Boyd and Chua [2.37] realised the importance and difficulty of Volterra kernel estimation and proposed a new method that allowed appropriate excitation design to avoid certain types of interference between different degree components. This way the kernel extraction procedure was simplified and the number of required measurements reduced. The basic idea advanced in this work was that, if a particularly sparse multisine is used, the output term frequency coincidences can be minimized and thus the number of kernel coefficients that can be obtained independently from a single measurement increases. However, this strategy does not solve the problem advanced before of different degree terms containing symmetric frequency pairs (ω,-ω,...) and sharing the same output frequency. A similar strategy was adopted nearly at the same time by Lawrence [2.41] whose aim was to design signals with autocorrelation properties identical to white noise up to a given order. While the solution proposed by Lawrence lead to a sparser signal, it also had the advantage of keeping a larger number of unique output

frequency combinations. In the 90's Evans *et al*. [2.42] compared these works and presented an interesting summary of the characteristics of each of them that takes into account their sparseness, auto-correlation properties, crest factor and the number of each order kernel points that can be measured directly. After the analyses they proposed a range of new periodic signals that present a good compromise between sparseness and output separability.

Orthogonal Approaches

A different way of optimize the Volterra coefficients' extraction is to make the matrix **X** in (2.26) diagonal. This way the numerical problems that can occur when trying to invert it are minimized. One possible way of do this is to rearrange the terms of the polynomial series so that the entries of the **X** matrix also change. In a recent work about orthogonal memoryless polynomial model extraction, a new general, non-recursive formula for the coefficients of a power series orthogonal to a particular uniformly distributed random variable input is presented [2.43]. In another work, the same author presents the formula for the computation of the polynomial coefficients in the case in which the input is Gaussian noise [2.44]. These works address a very important issue: the ill conditioning of the equation system used for the determination of the polynomial model coefficients. However the main limitation of this work is that it addresses this problem in a memoryless situation.

The work developed by Kim and Powers [2.45] is based on Gram-Schmidt orthogonalization of the different order input products used on the Volterra series. This orthogonalization leads to a set of modified orthogonal kernels that can be extracted by polyspectral correlations and later related to the original ones by the orthogonalization matrix. The main advantage of this work is the possibility of orthogonal extraction of the Volterra kernels independently of the input signal statistics (previous approaches required Gaussian input signals). This is an important contribution since not all the systems can be tested using Gaussian like signals. Additionally, sometimes this type of signals might not be very useful if they excite the system under test in a different operation point, when compared to the one obtained with the signals that the model is intended to process.

More recently, Cheng and Powers [2.46], presented an algorithm for the optimal determination (in mean square error sense) of the low pass equivalent Volterra series

coefficients for a nonlinear communication system processing PSK or QAM signals. Formal derivation of the cross-correlation formulas that minimize the MSE is presented for each type of fifth kernel (according to their multiplicity) when the input signal is a PSK and a QAM signal. In this work the dependence of the extracted model on the input signal considered is stated. Namely, it is shown that, in particular conditions, a third order Volterra model can be accurate enough to deal with a PSK signal, while for obtaining the same normalized mean square error (NMSE) with a QAM signal a fifth order Volterra model is required.

It is worth mentioning here a previously referred work [2.38] that also presents some contributions to the development of an orthogonal extraction procedure that allows the determination of Volterra kernel coefficients. In this work Gardner and Archer propose a method that allows the aimed orthogonal extraction using as input excitation cyclostacionary signals. However this is a class of complex input signals that can be used to extract the model of mathematically described nonlinear systems, but that cannot be applied in physical systems.

2.4 Conclusions

In this chapter some of the work that has been carried on by the scientific community on the behavioural modelling topic has been described. A brief description of the different phenomena observed in a PA was made, in order to show the effects that a behavioural model has to account for. It was shown that a lot of work has been done in behavioural modelling but its foundations are fragile.

The Volterra Series is a modelling topology suitable for behavioural modelling of PAs but whose coefficients are difficult to obtain with the existent standard extraction techniques. These difficulties can be minored if an appropriate extraction process is followed. Some previous existent work on VS coefficients orthogonal extraction was presented, despite some of these approaches are not well suited to extract the coefficients of a general system.

2.5 References

[2.1]. J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. on Microwave Theory & Tech.*, vol. MTT-53, Issue 1, pp. 1150-1163, April 2005.

[2.2]. V. Rizzoli, E. Montanari, D. Masotti, A. Lipparini and F. Mastri, "Domain-decomposition harmonic balance with block-wise constant spectrum," in *IEEE MTT-S Int. Microwave Symp. Digest*, San Francisco, June 2006, pp. 860 – 863.

[2.3]. V. Mathews and G. Sicuranza, *Polynomial Signal Processing,* New York: Wiley, 2000.

[2.4]. G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control Signals Systems*, vol. 2, pp. 303-314, 1989.

[2.5]. K. Hornik, M. Stinchcombe and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.

[2.6]. J. C. Pedro, N. B. Carvalho and P.M. Lavrador, "Modeling nonlinear behavior of bandpass memoryless and dynamic systems," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, U.S.A, June 2003, pp. 2133-2136.

[2.7]. P. M. Lavrador, J. C. Pedro and N. B. Carvalho, "Underlying linear system identification of memoryless and dynamic nonlinear systems," *4th National Conference on Telecommunications Digest,* Aveiro, Portugal 2003.

[2.8]. J. Vuolevi, T. Rahkonen, and J. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Trans. on Microwave Theory & Tech.*, vol. 49, pp. 1383-1389, 2001.

[2.9]. J. Vuolevi and T. Rahkonen, *Distortion in RF Power Amplifiers*, Norwood, MA: Artech House, 2003.

[2.10]. A. Saleh, "Frequency-indepedent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. on Commun.*, vol. 29, pp. 1715-1720, 1981.

[2.11]. N. Blachman, "Bandpass nonlinearities," *IEEE Trans. Inform. Theory*, vol. IT-10, pp. 162-164, 1964.

[2.12]. N. Blachman, "The Signal x Signal, Noise x Noise, and Signal x Noise output of a nonlinearity," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 21-27, 1968.

[2.13]. N. Blachman, "The effect of a large signal upon a small signal in a memoryless nonlinear bandpass amplifier," *IEEE Trans. on Commun.*, vol. 29, pp. 72-73, 1981.

[2.14]. H. Poza, Z. Sarkozy, and H. Berger, "A wideband data link computer simulation model," presented at NAECON, 1975.

[2.15]. M. Abuelma'atti, "Frequency-dependent nonlinear quadrature model for TWT amplifiers," *IEEE Trans. on Commun.*, vol. 32, no. 8, pp. 982-986, 1984.

[2.16]. C. Silva, A. Moulthrop, and M. Muha, "Introduction to polyspectral modeling and compensation techniques for wideband communications systems," *in 58th ARFTG Conference Digest*, San Diego, CA, Nov. 2001, pp 1-15.

[2.17]. M. Ibnkahla, N. J. Bershad, J. Sombrin, and F. Castanié, "Neural network modeling and identification of nonlinear channels with memory: Algorithms, applications, and analytic models," *IEEE Trans. Signal Process.*, vol. 46, pp. 1208-1220, 1998.

[2.18]. Y. Fang, M. C. Yagoub, F. Wang, and Q. J. Zhang, "A new macromodeling approach for nonlinear microwave circuits based on recurrent neural networks," *IEEE Trans. Microw. Theory Tech.*, vol. 48, pp. 2335–2344, 2000.

[2.19]. D. Mirri, F. Filicori, G. Iuculano, and G. Pasini, "A nonlinear dynamic model for performance analysis of large-signal amplifiers in communication systems," *in IMTC/99 IEEE Instrum. Meas. Technol. Conference Digest*, Venice, Italy, May 1999, pp193-197.

[2.20]. E. Ngoya, N. L. Gallou, J. M. Nébus, H. Burêt, and P. Reig, "Accurate RF and microwave system level modeling of wide band nonlinear circuits," in *IEEE MTT-S Int. Microwave Symp. Digest*, Boston, MA, June 2000, pp. 79-82.

[2.21]. A. Soury, E. Ngoya, and J. M. Nebus, "A new behavioral model taking into account nonlinear memory effects and transient behaviors in wideband SSPAs," in *IEEE MTT-S Int. Microwave Symp. Digest*, Seattle, WA, June 2002, pp. 853-856.

[2.22]. F. Filicori, G. Vannini, and V. Monaco, "A nonlinear integral model of electron devices for HB circuit analysis," *IEEE Trans. Microw. Theory Tech.*, vol. 40, pp. 1456-1465, 1992.

[2.23]. H. Ku, M. Mckinley, and J. S. Kenney, "Quantifying memory effects in RF power amplifiers," *IEEE Trans. Microw. Theory Tech.*, vol. 50, pp. 2843-2849, 2002.

[2.24]. P. Asbeck, H. Kobayashi, M. Iwamoto, G. Hanington, S. Nam, and L. E. Larson, "Augmented behavioral characterization for modeling the nonlinear response of power amplifiers," in *IEEE MTT-S Int. Microwave Symp. Digest*, Seattle, WA, June 2002, pp. 135-138.

[2.25]. P. Draxler, I. Langmore, T. P. Hung, and P. M. Asbeck, "Time domain characterization of power amplifiers with memory effects," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, PA, June 2003, pp. 803-806.

[2.26]. J. Dooley, B. O'Brien, and T. J. Brazil, "Behavioral modeling of RF power amplifiers using adaptive recursive polynomial functions," in *IEEE MTT-S Int. Microwave Symp. Digest*, San Francisco, June 2006, pp. 852-855.

[2.27]. T.Wang and T. J. Brazil, "Volterra-mapping-based behavioral modeling of nonlinear circuits and systems for high frequencies," *IEEE Trans. Microw. Theory Tech.*, vol. 51, pp. 1433-1440, 2003.

[2.28]. A. Zhu, M. Wren, and T. Brazil, "An efficient Volterra-based behavioral model for wideband RF power amplifiers," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, PA, June 2003, pp 787-790.

[2.29]. A. Zhu, J. Dooley, and T. J. Brazil, "Simplified Volterra series based behavioral modeling of RF power amplifiers using deviation-reduction," in *IEEE MTT-S Int. Microwave Symp. Digest*, San Francisco, June 2006, pp. 1113-1116.

[2.30]. A. Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned volterra series," *IEEE Microwave and WirelessComponents Letters*, vol. 14, pp. 563-565, 2004.

[2.31]. M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: Wiley, 1980.

[2.32]. W. J. Rugh, *Nonlinear System Theory—The Volterra–Wiener Approach*. Baltimore, MD: The Johns Hopkins Univ. Press, 1981.

[2.33]. M. Isaksson, D. Wisell, and D. Ronnow, "A comparative analysis of behavioral models for RF power amplifiers," *IEEE Trans. on Microwave Theory & Tech.*, vol. 54, pp. 348-359, 2006.

[2.34]. M. Isaksson and D. Rönnow, "A Kautz-Volterra Behavioral Model for RF Power Amplifiers," in *IEEE MTT-S Int. Microwave Symp. Digest*, San Francisco, June 2006, pp. 485-488.

[2.35]. A. Zhu and T. J. Brazil, "RF power amplifier behavioral modeling using Volterra expansion with Laguerre functions," in *IEEE MTT-S Int. Microwave Symp. Digest*, June 2005, pp. 963-966.

[2.36]. P. Koukoulas and N. Kalouptsidis, "Nonlinear system identification using gaussian inputs," *IEEE Trans. on Signal Processing*, vol. 43, pp. 1831-1841, 1995.

[2.37]. S. Boyd, Y. S. Tang, and L. O. Chua, "Measuring volterra kernels," *IEEE Trans. on Circuits and Systems*, vol. CAS-30 pp. 571-577, 1983.

[2.38]. W. A. Gardner and T. L. Archer, "Exploitation of cyclostationarity for identifying the Volterra kernels of nonlinear systems," *IEEE Trans. Inform. Theory*, vol. 39, pp. 535-542, 1993.

[2.39]. R. Pintelon and J. Schoukens, *System Identification, A Frequency Domain Approach*. Piscataway, NJ: IEEE, 2001.

[2.40]. K. Vanhoenacker, T. Dobrowiecki, and J. Schoukens, "Design of multisine excitations to characterize the nonlinear distortions during FRF-measurements," *IEEE Trans. on Instrumentation and Measurement*, vol. 50, pp. 1097-1102, 2001.

[2.41]. P. J. Lawrence, "Estimation of the Volterra functional series of a nonlinear system using frequency response data," *IEE Proc. D*, vol. 128, pp. 206-210, 1981.

[2.42]. C. Evans, D. Rees, L. Jones, and M. Weiss, "Periodic signals for measuring nonlinear Volterra kernels," *IEEE Trans. on Instrumentation and Measurement*, vol. 45, pp. 362-371, 1996.

[2.43]. R. Raich, H. Qian, and G. T. Zhou, "Orthogonal polynomial for power amplifier modelling and predistorter design," *IEEE Trans. on Vehicular Technology*, vol. 53, pp. 1468-1479, 2004.

[2.44]. R. Raich and G. T. Zhou, "Orthogonal polynomials for complex gaussian processes," *IEEE Trans. on Signal Processing*, vol. 52, pp. 2788-2797, 2004.

[2.45]. S. B. Kim and E. J. Powers, "Orthogonalised frequency domain Volterra model for non-Gaussian inputs," *IEE proceedings-F*, vol. 140, pp. 402-409, 1993.

[2.46]. C. H. Cheng and E. J. Powers, "Optimal Volterra kernel estimation algorithms for a nonlinear communication system for PSK and QAM inputs," *IEEE Trans. on Signal Processing*, vol. 49, pp. 147-163, 2001.

3. The Orthogonal Formulation for Volterra Series Extraction

3.1 Introduction

In the previous chapter the state of art in power amplifier behavioural modeling has been presented and the need for an improved method of Volterra series coefficient determination was introduced in order to improve model accuracy and range of validity.

In this chapter the whole process leading to the formulation of the orthogonal polynomial used to obtain the Volterra series description of a nonlinear system is presented. The model topology and theoretical support are described in section 3.2. Sections 3.3 and 3.4 describe the input signal selection and the assembly of each order terms to build the orthogonal model. Section 3.5 presents an analysis of the orthogonal coefficient determination convergence process, while section 3.6 shows how to convert the orthogonal coefficients back into the time domain Volterra kernels. Finally section 3.7 summarizes and concludes this chapter.

3.2 The Used Model Topology

The modeling approach herein presented is founded on the rigorous nonlinear system identification theory. This states that any single-input / single-output nonlinear dynamic system that is stable and of fading memory, can be represented by a cascade of a single-input multiple-output linear system with memory, followed by a multiple-input single-output nonlinear memoryless system [3.1,3.2]. One possible implementation of this is the nonlinear finite impulse response filter stated in (3.1).

$$y(s) = f_{NL}[x(s), x(s-1),..., x(s-M)] \tag{3.1}$$

In this expression M, indicates the number of time delays considered (the system's memory span or depth) while $f_{NL}(.)$ is any (M+1) to 1 nonlinear universal approximator. Two widely used implementations of this universal approximator are the artificial neural network (ANN) – which leads to the time-delay ANN nonlinear filter, and the multidimensional polynomial – leading to the general polynomial filter (PF) or Volterra filter [3.3].

Both of these approaches present advantages and drawbacks. For example, while the relation between the ANNs' response to their parameter set is nonlinear, PFs are linear in their parameters. So, these latter models can be identified using direct extraction procedures, while ANNs require a nonlinear optimization process (of non guaranteed success in terms of convergence and uniqueness of the parameter set). Unfortunately, although the extraction of the polynomial model parameter set can be performed solving a linear system of equations, this is not easy in practice, due to the large number of coefficients to be determined, and the fact that the poor separability of some equations tends to produce an ill-conditioned system. Usually, such a problem is circumvented through the use of some optimization process, for instance, via an adaptation loop. But that jeopardizes the advantage of direct extraction.

In this work we intend to profit from the parameter linearity of the PFs and circumvent the problem of ill-conditioning by rearranging the Volterra series in order to obtain an orthogonal model. This work is similar to the one made by [3.1], except that now instead

of considering white Gaussian input signal a set of randomized phase multisines are used.

3.3 Input Signal Selection to Build an Orthogonal Polynomial

One must realize that the orthogonality condition of a certain polynomial will hold only for one particular input signal. So, it is of major importance to choose the particular input signal for which the polynomial should be orthogonal. This signal must verify two different conditions: (i) it should guarantee that all the system states are excited, so that a general complete model is extracted, and (ii) it should be easy to synthesize in a RF laboratory allowing this way a simple feasible model extraction procedure. A signal known to meet both of these conditions is a set of multisines with randomized phases. Actually it is known that this signal is a complete signal, that is, is able to excite all different system states [3.4], and it is easy to generate and measure in a usual RF laboratory.

Since it was already determined that a Volterra series approximator will be used to describe $f_{NL}[.]$ in (3.1), it can be re-written as:

$$\begin{aligned} y(s) \;=\; & h_0 \\ & + \sum_{m=0}^{M-1} h_m x(s-m) \\ & + \sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{m_1 m_2} x(s-m_1)x(s-m_2) \\ & + \ldots \\ & + \sum_{m_1=0}^{M-1}\ldots\sum_{m_N=0}^{M-1} h_{m_1\ldots m_N} x(s-m_1)\ldots x(s-m_N) \end{aligned} \tag{3.2}$$

where M is the memory span and N the order of the nonlinearity.

To achieve the orthogonal coefficients extraction it is mandatory to rearrange the terms in (3.2) so that all of them become separable for the particular input of a set of multi-sines with randomized phases.

Prior to state this process in a formal way, lets look at Fig. 3.1 that presents the final orthogonal model topology in a schematic way.

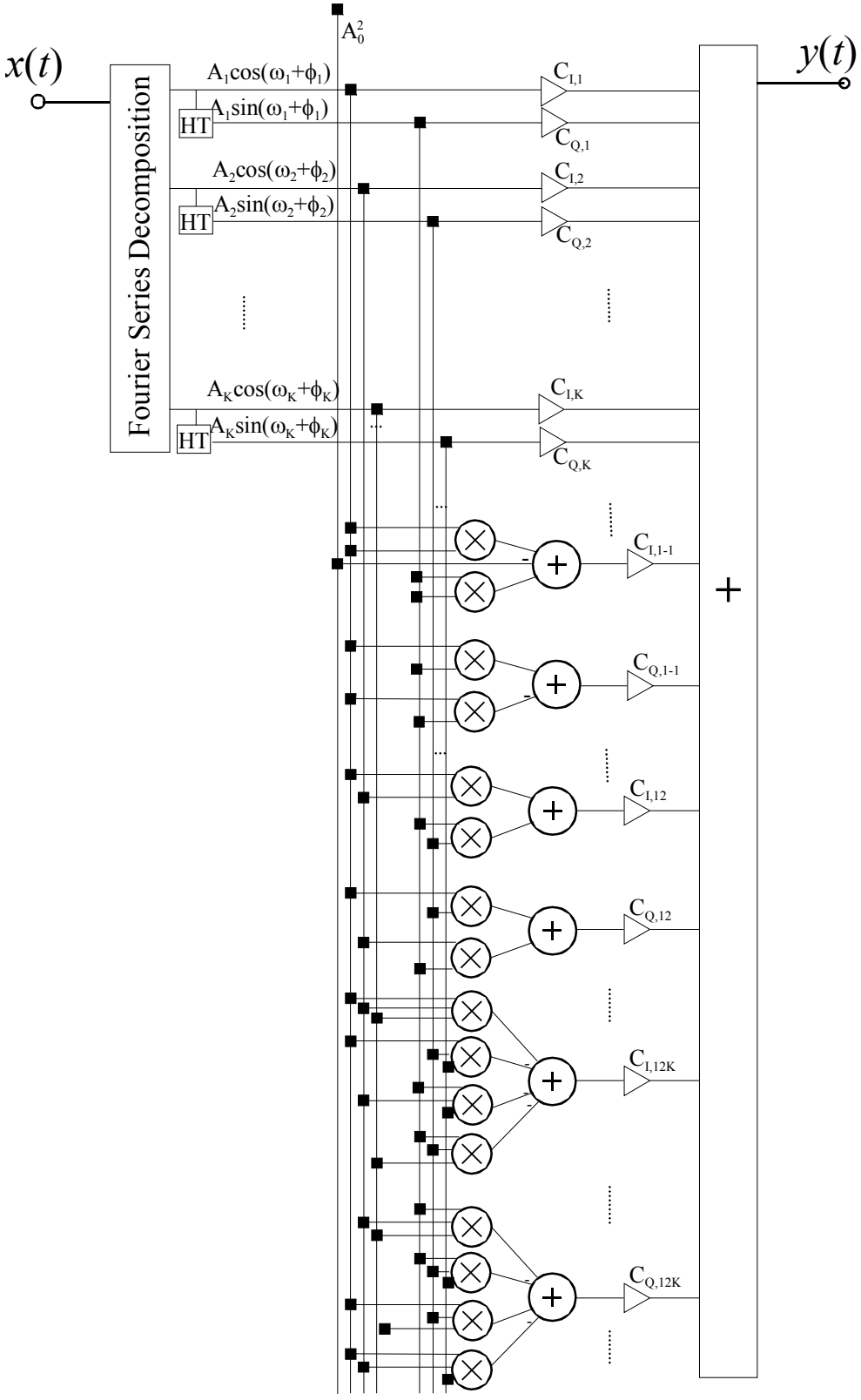


Figure 3.1 – Schematic representation of the dynamic polynomial model topology used.

In Figure 3.1, it is seen the Fourier decomposition of the input signal into a set of single

tone signals. Those signals will then be processed by nonlinear combinations arranged so that each of the combinations produces a separable output component. By guaranteeing that the output of each block is separable from all the others we obtain our orthogonal model.

Note that this model verifies the topology proposed by [3.1,3.2] of a linear dynamic single input / multiple output system – the Fourier series decomposition – and a static multiple input / single output nonlinear system.

In order to write this in a formal way the statement of the Fourier series decomposition used is stated:

$$x(s) = \sum_{k=0}^{K-1} C_k \cos(\omega_k s) + S_k \sin(\omega_k s). \tag{3.3}$$

in which the C_k and S_k coefficients are given by:

$$C_k = \frac{1}{K} \sum_{s=0}^{K-1} x(s) \cos(\omega_k s) \tag{3.4a}$$

and

$$S_k = \frac{1}{K} \sum_{s=0}^{K-1} x(s) \sin(\omega_k s) \tag{3.4b}$$

where

$$\omega_k = \frac{2\pi}{K} k \tag{3.5}$$

Now, all *n*'th order responses of (3.2) should be given as a function of these frequency domain inputs. For example, for the first order, or linear part, of (3.2) the corresponding output is:

$$y_1(s) = \sum_{m=0}^{M-1} h_m \sum_{k=0}^{K-1} C_k \cos[\omega_k (s - m)] + S_k \sin[\omega_k (s - m)] \tag{3.6}$$

Separating the cosine and sine dependence on *s* and *m*, grouping the terms in *m* and then changing the summation order:

$$y_1(s) = \sum_{k=0}^{K-1} [C_k \cos(\omega_k s) + S_k \sin(\omega_k s)] \sum_{m=0}^{M-1} h_m \cos(\omega_k m) + [C_k \sin(\omega_k s) - S_k \cos(\omega_k s)] \sum_{m=0}^{M-1} h_m \sin(\omega_k m) \quad (3.7)$$

This is exactly the required frequency domain representation that is being looked for. Grouping the terms in this expression as:

$$\begin{aligned} x_k(s) &= [C_k \cos(\omega_k s) + S_k \sin(\omega_k s)] \\ x_k'(s) &= C_k \sin(\omega_k s) - S_k \cos(\omega_k s) \\ H_k^C &= \sum_{m=0}^{M-1} h_m \cos(\omega_k m) \\ H_k^S &= \sum_{m=0}^{M-1} h_m \sin(\omega_k m) \end{aligned} \quad (3.8)$$

where $x_k'(s)$ indicates a 90º phase shift, and H_k^C and H_k^S are the frequency domain coefficients of the linear transfer function. Equation (3.7) can be rewritten in compact form as:

$$y_1(s) = \sum_{k=0}^{K-1} H_k^C x_k(s) + H_k^S x_k'(s) \quad (3.9)$$

A similar process can be applied to the second order part of (3.2) using the bi-dimensional form of the Fourier series. Once again the starting point is the second order time domain system's response expression:

$$y_2(s) = \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} h_{m_1 m_2} x(s - m_1) x(s - m_2) \quad (3.10)$$

writing $x(s)$ in its Fourier series:

$$y_2(s) = \sum_{m_1=0}^{M-1} \sum_{m_2=0}^{M-1} h_{m_1 m_2} \left\{ \sum_{k_1=0}^{K-1} \left[C_{k_1} \cos\left(\omega_{k_1}(s - m_1)\right) + S_{k_1} \sin\left(\omega_{k_1}(s - m_1)\right) \right] \cdot \sum_{k_2=0}^{K-1} \left[C_{k_2} \cos\left(\omega_{k_2}(s - m_2)\right) + S_{k_2} \sin\left(\omega_{k_2}(s - m_2)\right) \right] \right\} \quad (3.11)$$

separating s from m in each sine and cosine and rearranging the expression:

$$\begin{aligned} y_2(s) = \sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1}\sum_{k_1=0}^{K-1}\sum_{k_2=0}^{K-1} \; & \Big\{h_{m_1m_2}\Big[\big(C_{k_1}\cos(\omega_{k_1}s) + S_{k_1}\sin(\omega_{k_1}s)\big)\cos(\omega_{k_1}m_1) + \\ & \big(C_{k_1}\sin(\omega_{k_1}s) - S_{k_1}\cos(\omega_{k_1}s)\big)\sin(\omega_{k_1}m_1)\Big] \\ & + \Big[\big(C_{k_2}\cos(\omega_{k_2}s) + S_{k_2}\sin(\omega_{k_2}s)\big)\cos(\omega_{k_2}m_2) \\ & \big(C_{k_2}\sin(\omega_{k_2}s) - S_{k_2}\cos(\omega_{k_2}s)\big)\sin(\omega_{k_2}m_2)\Big]\Big\} \end{aligned} \tag{3.12}$$

Changing the summation orders and simplifying the notation of $x_i(s)$

$$\begin{aligned} y_2(s) = \sum_{k_1=0}^{K-1}\sum_{k_2=0}^{K-1}\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{m_1m_2}\Big[& x_{k_1}(s)x_{k_2}(s)\cos(\omega_{k_1}m_1)\cos(\omega_{k_2}m_2) \\ & + x_{k_1}(s)x'_{k_2}(s)\cos(\omega_{k_1}m_1)\sin(\omega_{k_2}m_2) \\ & + x'_{k_1}(s)x_{k_2}(s)\sin(\omega_{k_1}m_1)\cos(\omega_{k_2}m_2) \\ & + x'_{k_1}(s)x'_{k_2}(s)\sin(\omega_{k_2}m_2)\sin(\omega_{k_2}m_2)\Big] \end{aligned} \tag{3.13}$$

according to the definition of the bi-dimensional coefficients and splitting each h_{m1m2} into different parts we get

$$\begin{aligned} y_2(s) = \sum_{k_1=0}^{K-1}\sum_{k_2=0}^{K-1}\Bigg[& x_{k_1}(s)x_{k_2}(s)\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{m_1m_2}\cos(\omega_{k_1}m_1)\cos(\omega_{k_2}m_2) \\ & + x_{k_1}(s)x'_{k_2}(s)\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{m_1m_2}\cos(\omega_{k_1}m_1)\sin(\omega_{k_2}m_2) \\ & + x'_{k_1}(s)x_{k_2}(s)\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{m_1m_2}\sin(\omega_{k_1}m_1)\cos(\omega_{k_2}m_2) \\ & + x'_{k_1}(s)x'_{k_2}(s)\sum_{m_1=0}^{M-1}\sum_{m_2=0}^{M-1} h_{m_1m_2}\sin(\omega_{k_2}m_2)\sin(\omega_{k_2}m_2)\Bigg] \end{aligned} \tag{3.14}.$$

in which each of the double summations in m_1 and m_2 are part of the bi-dimensional Fourier series. So (3.14) can be re-written as:

$$y_2(s) = \sum_{k_1=0}^{K-1}\sum_{k_2=0}^{K-1}\Big[x_{k_1}(s)x_{k_2}(s)H^{CC}_{k_1k_2} + x_{k_1}(s)x'_{k_2}(s)H^{CS}_{k_1k_2} + x'_{k_1}(s)x_{k_2}(s)H^{SC}_{k_1k_2} + x'_{k_1}(s)x'_{k_2}(s)H^{SS}_{k_1k_2}\Big] \tag{3.15}$$

This is the desired expression: the second order output written as a function of the input Fourier components and frequency domain coefficients.

This process can also be applied to the *n*'th order component of (3.2) using the *n*-dimensional form of the Fourier series. In this general case:

$$y_N(s) = \sum_{k_1=0}^{K-1} \cdots \sum_{k_N=0}^{K-1} \left[x_{k_1}(s) \ldots x_{k_N}(s) H_{k_1 \ldots k_N}^{C \ldots C} + x_{k_1}(s) \ldots x'_{k_N}(s) H_{k_1 \ldots k_N}^{C \ldots S} + \right. \\ \left. + x'_{k_1}(s) \ldots x_{k_N}(s) H_{k_1 \ldots k_N}^{S \ldots C} + x'_{k_1}(s) \ldots x'_{k_N}(s) H_{k_1 \ldots k_N}^{S \ldots S} \right] \tag{3.16}$$

After this derivation a model formulation that approximates the system's response using each of the input signal's frequency components was obtained. Now, a rearrangement of its terms to get a model that is orthogonal for a multi-sine at a given input amplitude is required.

3.4 Obtaining the Model Orthogonality

In order to build an orthogonal model an inner product definition is required. This will then be used to express the corresponding orthogonality condition, which the model components, Ψ, must fulfill. Since the input signal for orthogonality are randomized phase multi-sines, the inner product is the average of a large number (*R*, ideally infinite) of different phase arrangement realizations, *r*:

$$\left\langle \Psi_{nI}\left[x_{k_1}(s) \ldots x_{k_n}(s)\right] \bullet \Psi_{mI}\left[x_{k_1}(s) \ldots x_{k_m}(s)\right] \right\rangle \\ = \lim_{R \to \infty} \frac{1}{R} \sum_{r=1}^{R} \sum_{s=0}^{K-1} \Psi_{nI}\left[x_{k_1}(s) \ldots x_{k_n}(s)\right] \Psi_{mI}\left[x_{k_1}(s) \ldots x_{k_m}(s)\right] \tag{3.17}$$

where Ψ_{nI} represents an individual term of the cosine part of order *n,* of the orthogonal polynomial that we are looking for.

It can be shown that (3.17) verifies the properties of a general inner product. Therefore, the orthogonal model formulation requires the development of a complete set of model basis functions, Ψ_n and Ψ_m, where the only non null inner products of (3.17) are the ones in which *n*=*m*. As is known from Fourier series, the inner product of terms at different frequencies is zero. So, for that type of model functions the orthogonality is already guaranteed. However, since in this case the input signal is an evenly spaced multi-sine,

there will appear some Ψ_x's that share the same output frequency, and these require special attention to guarantee orthogonality.

There are two different things that may cause two output components to share the same frequency: (*i*) different frequency combinations of evenly distributed tones (e.g. $\omega_1+\omega_2-\omega_3$ and $\omega_1+\omega_3-\omega_4$) and (*ii*) interference between components of different orders (e.g. ω_1 and $\omega_1+\omega_1-\omega_1$ or $\omega_1+\omega_2-\omega_2$).

These two situations are quite different. Type (*i*) interfering terms share the same frequency, but their phases vary independently. Therefore, each of the contributions can be separated from the others, when the average of the responses to multi-sines of randomized phases is performed [along with *r*, in (3.17)].

Type (*ii*) interfering terms can not be overcome this way since the phase of those contributions is exactly the same. So, the cancellation of type (*ii*) interference will be achieved by adding orthogonalization parts to those terms. As the desired orthogonal extraction stimulus is a multi-sine in which all the tones have A_0 amplitude, only the lowest order of type *ii*) terms will be kept, while all the others will be canceled by subtracting the value of their responses at the extraction level. This way in the example of first and third order interference, the first order coefficient accounts for the linear and non-linear correlated part [3.5].

So, the orthogonal model is achieved by grouping, in a systematic way, all the input combinations to obtain model terms that produce a single output frequency, and by adding special orthogonalization parts to some of the terms that share the same frequency.

Basically, the resulting model is a generalization of the Chebyshev polynomial (defined for one tone input and for memoryless systems), into a general polynomial with memory that is now orthogonal for the *K* input tones.

After all these considerations the general orthogonal model creation procedure is stated. The first function considered is the one that produces the dc term. This term has no phase and so the quadrature function is zero valued. The result for the zero order components is:

$$\Psi_{0I} = 1 \tag{3.18a}$$

and

$$\Psi_{0Q} = 0\,. \tag{3.18b}$$

The first order components of the model are linearly scaled versions of the inputs. So the model functions are $x_k(s)$ and $x_k'(s)$, the output of the linear decomposition:

$$\Psi_{1I}(\omega_k, s) = A_k \cos(\omega_k s + \phi_k) = C_k \cos(\omega_k s) + S_k \sin(\omega_k s) \equiv x_k \quad (3.19a)$$

[1]

and

$$\Psi_{1Q}(\omega_k, s) = A_k \sin(\omega_k s + \phi_k) = S_k \cos(\omega_k s) - C_k \sin(\omega_k s) \equiv x_k'. \quad (3.19b)$$

As they are linear, they produce a single output and thus need no orthogonalization. Note that x_k is a function of the time sample s, $x_k(s)$, but, for simplicity of notation, its dependence on (s) is omitted.

The second order functions can produce the frequencies $\omega_i+\omega_j$ or $\omega_i-\omega_j$ so there should be an independent Ψ for the sum and another for difference frequencies. If $i=j$ then the second order difference contribution will fall at dc. The dc component was already accounted for in the zero order. So, the second order function can not produce another dc output, when the input amplitude is A_0 [that is the role of the delta function in (3.20a)]. The contribution of input signal amplitude to the change of the output dc level is accounted for at a different input amplitude level. The formulation of these functions is

$$\Psi_{2I}(\omega_{k_1} \pm \omega_{k_2}, s) = x_{k_1} x_{k_2} \mp x_{k_1}' x_{k_2}' - \delta(\omega_{k_1} \pm \omega_{k_2}) A_0^2 \quad (3.20a)$$

and

$$\Psi_{2Q}(\omega_{k_1} \pm \omega_{k_2}, s) = x_{k_1}' x_{k_2} \pm x_{k_1} x_{k_2}' \quad (3.20b)$$

The third order functions can produce the frequencies $\pm \omega_i \pm \omega_j \pm \omega_k$. If two symmetric frequencies appear in a combination (e.g $\omega_j=-\omega_k$) the resultant frequency will be coincident with the first order component (ω_i). As was stated above, the third order function must not produce an output coincident with ω_i, when the input amplitude is A_0, to enable the separate measurement of the linear coefficient. The third order impact at that frequency will be measured later. The functions obtained are:

[1] Note that the $\Psi(\omega_k, s)$ is a function of the input component $x_k(s)$ whose resulting frequency is at ω_k.

$$\begin{aligned}\Psi_{3I}(\omega_{k_1}+\omega_{k_2}+\omega_{k_3},s) = {} & \Psi_{1I}(\omega_{k_1},s)\Psi_{1I}(\omega_{k_2},s)\Psi_{1I}(\omega_{k_3},s)\\ & -\Psi_{1I}(\omega_{k_1},s)\Psi_{1Q}(\omega_{k_2},s)\Psi_{1Q}(\omega_{k_3},s)\\ & -\Psi_{1Q}(\omega_{k_1},s)\Psi_{1I}(\omega_{k_2},s)\Psi_{1Q}(\omega_{k_3},s)\\ & -\Psi_{1Q}(\omega_{k_1},s)\Psi_{1Q}(\omega_{k_2},s)\Psi_{1I}(\omega_{k_3},s)\\ & -\delta(\omega_{k_1}+\omega_{k_2})A_0^2\Psi_{1I}(\omega_{k_3},s)\\ & -(1-\delta(\omega_{k_2}-\omega_{k_3}))\delta(\omega_{k_1}+\omega_{k_3})A_0^2\Psi_{1I}(\omega_{k_2},s)\\ & -(1-\delta(\omega_{k_1}-\omega_{k_2}))\delta(\omega_{k_2}+\omega_{k_3})A_0^2\Psi_{1I}(\omega_{k_1},s)\end{aligned} \tag{3.21a}$$

$$\begin{aligned}\Psi_{3Q}(\omega_{k_1}+\omega_{k_2}+\omega_{k_3},s) = {} & -\Psi_{1Q}(\omega_{k_1},s)\Psi_{1Q}(\omega_{k_2},s)\Psi_{1Q}(\omega_{k_3},s)\\ & -\Psi_{1I}(\omega_{k_1},s)\Psi_{1I}(\omega_{k_2},s)\Psi_{1Q}(\omega_{k_3},s)\\ & -\Psi_{1I}(\omega_{k_1},s)\Psi_{1Q}(\omega_{k_2},s)\Psi_{1I}(\omega_{k_3},s)\\ & +\Psi_{1Q}(\omega_{k_1},s)\Psi_{1I}(\omega_{k_2},s)\Psi_{1I}(\omega_{k_3},s)\\ & -\delta(\omega_{k_1}+\omega_{k_2})A_0^2\Psi_{1Q}(\omega_{k_3},s)\\ & -(1-\delta(\omega_{k_2}-\omega_{k_3}))\delta(\omega_{k_1}+\omega_{k_3})A_0^2\Psi_{1Q}(\omega_{k_2},s)\\ & -(1-\delta(\omega_{k_1}-\omega_{k_2}))\delta(\omega_{k_2}+\omega_{k_3})A_0^2\Psi_{1Q}(\omega_{k_1},s)\end{aligned} \tag{3.21b}$$

For simplicity of notation, in these expressions it was assumed that for components involving frequency differences the minus signal is hidden inside the ω_k. This simplification lacks physical meaning, since there are no negative frequencies. But, remembering that $\Psi_{1I}(\omega_k,s)$ is a cosine of ω_k and that $\Psi_{1Q}(\omega_k,s)$ is the sine part, then the assumption made corresponds to change the signal of $\Psi_{1Q}(\omega_k,s)$ when the term desired involves $-\omega_k$.

These expressions obtained can be generalized recursively obtaining the result

$$\begin{aligned}\Psi_{NI}(\omega_1+\ldots+\omega_N,s) = {} & \Psi_{(N-1)I}(\omega_1+\ldots+\omega_{N-1},s)\Psi_{1I}(\omega_N,s)\\ & -\Psi_{(N-1)Q}(\omega_1+\ldots+\omega_{N-1},s)\Psi_{1Q}(\omega_N,s)\\ -A_0^2\sum_{n=1}^{N-1}\sum_{k_1=1}^{N-1}\ldots\sum_{k_n=k_{n-1}+1}^{N-1} & \{\delta(\omega_{k_1}+\omega_N)..\delta(\omega_{k_n}+\omega_N)\cdot\\ & \cdot\prod_{m_1=1}^{N-1}\ldots\prod_{m_n=m_{n-1}+1}^{N-1}(1-\delta(\omega_{m_1}-\omega_k)..\delta(\omega_{m_n}-\omega_k))\Psi_{(N-2)I}(\omega_1+\ldots+\omega_{k-1}+\omega_{k+1}+\ldots+\omega_{N-1},s)\}\end{aligned} \tag{3.22a}$$

$$\Psi_{NQ}(\omega_1 + ... + \omega_N, s) = \Psi_{(N-1)I}(\omega_1 + ... + \omega_{N-1}, s)\Psi_{1Q}(\omega_N, s) + \Psi_{(N-1)Q}(\omega_1 + ... + \omega_{N-1}, s)\Psi_{1I}(\omega_N, s)$$
$$- A_0^2 \sum_{n=1}^{N-1} \sum_{k_1=1}^{N-1} ... \sum_{k_n=k_{n-1}+1}^{N-1} \left\{ \delta(\omega_{k_1} + \omega_N)..\delta(\omega_{k_n} + \omega_N) \cdot \right.$$
$$\left. \cdot \prod_{m_1=1}^{N-1} ... \prod_{m_n=m_{n-1}+1}^{N-1} (1 - \delta(\omega_{m_1} - \omega_k)..\delta(\omega_{m_n} - \omega_k)) \Psi_{(N-2)Q}(\omega_1 + ... + \omega_{k-1} + \omega_{k+1} + ... + \omega_{N-1}, s) \right\} \quad (3.22b)$$

Basically, the first line of (3.22a) and (3.22b) produces the resultant frequency from the combination (ω_1,…,ω_n), based on the frequency sets ω_n and (ω_1,…,ω_{n-1}). The summations and products at the second line evaluate the need of inclusion of an extra orthogonalization factor if ω_n is the symmetric of any of the frequencies in the range (ω_1,…,ω_{n-1}).

All the model terms presented are multiplied by scaling coefficients C_I and C_Q, the parameters of any instantiation of the model as was depicted in Figure 3.1, and as stated in (3.23).

$$y(s) = \sum_{n=0}^{N} \sum_{m_1=0}^{M-1} ... \sum_{m_n=m_{n-1}}^{M-1} C_{nI,m_1,...m_n} \psi_{nI}(\omega_{m_1} + ... + \omega_{m_n}, s) + C_{nQ,m_1,...m_n} \psi_{nQ}(\omega_{m_1} + ... + \omega_{m_n}, s) \quad (3.23)$$

3.5 Orthogonal Model Coefficients' Extraction

The model coefficients' extraction starts with the system excitation with a set of random phase multi-sines

$$x_r(s) = \sum_{n=0}^{N-1} A_0 \cos((\omega_0 + n\Delta\omega)s + \phi_{r,n}) \quad (3.24)$$

in which each tone has amplitude A_0, the defined amplitude of orthogonality, and the output components at each frequency are measured in amplitude and phase. The value of each coefficient can be calculated as the projection, based on the inner product (3.17), of the output over the corresponding orthogonal model component.

The coefficient determination process is based on averaging several random components in which each of them contains the desired coefficient and also a random

perturbation (from other components that share the same spectral position). So it is considered that the extraction process is finished after a certain number of random experiments that guarantee that the expected value of the interference on each coefficient is as small as prescribed. There are some strategies that can speed up the convergence. For example, if we start with the estimation of the first order coefficients and then use the obtained values for the estimation of the higher order ones, we will obtain better estimates of higher order coefficients with a smaller number of random phases in the input multi-sine.

In order to quantify this impact and determine what should be a reasonable number of random phases to consider, lets consider, as a simple example, a three tone multi-sine of equally spaced frequencies ω_1, $\omega_2 = \omega_1 + \Delta\omega$ and $\omega_3 = \omega_1 + 2\Delta\omega$. The output at frequency ω_1 will have two different sources. One resulting from the first order basis function (3.18a) and another of the third order basis function here rewritten:

$$\Psi_{3I}\left[x_2(s),\, x_2'(s),\, x_2(s),\, x_2'(s),\, x_3(s),\, x_3'(s)\right] = A_2^2 A_3 \cos((2\omega_2 - \omega_3)s + 2\phi_2 - \phi_3) \tag{3.25}$$

Because the model is orthogonal to this input, the coefficients (model kernels) multiplying those functions can be recursively determined using the projections of the output over the corresponding orthogonal basis functions. This projection is given by the ratio between (i) the output and the basis function cross-correlation and (ii) the corresponding basis function auto-correlation. So, in the first parameter extraction stage, the first order coefficients, $C_{1I,1}$, are estimated from

$$C_{1I,1} = \frac{\frac{1}{R}\sum_{r=1}^{R}\frac{1}{S}\sum_{s=1}^{S} y_r(s)\Psi_{1I}[x_{1r}(s)]}{\frac{1}{R}\sum_{r=1}^{R}\frac{1}{S}\sum_{s=1}^{S} \Psi_{1I}[x_{1r}(s)]\Psi_{1I}[x_{1r}(s)]} \tag{3.26}$$

In which R is the number of random phase measurement made and S the number of the time samples s. The numerator of (3.26) can always be understood as one part that relates $\Psi_{1I}[x_1(s)]$ with the component of $y(s)$ of the same phase and another part that relates $\Psi_{1I}[x_1(s)]$ with the part of $y(s)$ that has random phase. As seen from this numerator, the first part is the one that we are interested to extract, while the second one is the interference to

the measurement. Note that, because the various multi-sine realizations have randomized phases, the mentioned cross-correlation (in which the average is achieved during the summation of all the multi-sine realizations, r) guarantees that the interference indeed goes asymptotically to zero.

Then, using these first order coefficients, the $C_{3I,22\text{-}3}$ can be estimated from (3.27) in which the strong interference of the linear components was previously suppressed. In fact, since the extracted $C_{1I,1}$ will always have a residual error, there is still some interference from first to third order components that must be eliminated by the appropriate cross-correlation of (3.27). In a general situation (with more than three input frequencies) different third order combinations would appear at the same output frequency and thus should be also accounted here as interference.

$$C_{3I,22-3} = \sum_{r=1}^{R} \frac{\sum_{s=1}^{S} \{y_r(s) - C_{1I,1}\Psi_{1I}[x_{1r}(s)]\}\Psi_{3I}[x_{2r}(s), x'_{2r}(s), x_{2r}(s), x'_2 r(s), x_{3r}(s), x'_{3r}(s)]}{\sum_{s=1}^{S} \Psi_{3I}[x_{2r}(s), x'_{2r}(s), x_{2r}(s), x'_{2r}(s), x_{3r}(s), x'_{3r}(s)]\Psi_{3I}[x_{2r}(s), x'_{2r}(s), x_{2r}(s), x'_{2r}(s), x_{3r}(s), x'_{3r}(s)]} \tag{3.27}$$

Simplifying expression (3.27), for this particular case of a three tone input, and focusing only on the interference term, the following result for the error is reached:

$$Error(R) = \frac{\varepsilon_1 A_1}{A_2^2 A_3 N} \sum_{r=1}^{R} \cos(\phi_{1r} - 2\phi_{2r} + \phi_{3r}) \tag{3.28}$$

where ε_1 is the error limit for the estimation of $C_{1I,1}$. Since the phases ϕ are stochastic variables, the sum ϕ_1-$2\phi_2$+ϕ_3 is also a random variable. Using the central limit theorem, it can be shown that the summation of the random phase cosines becomes a Gaussian variable whose variance is $1/(2R)$ [3.6].

The problem now is to determine the value R that ensures, with a probability level P_L, that the error obtained is smaller than ε, that is:

$$P\left[\left|\frac{\varepsilon_1 A_1}{A_2^2 A_3 R} \sum_{r=1}^{R} \cos(\phi_{1r} - 2\phi_{2r} + \phi_{3r})\right| < \varepsilon\right] > P_L \tag{3.29}$$

Since the error is a Gaussian variable, this condition can be equivalently written as:

$$Q\left(\frac{\sqrt{2R}A_2^2A_3\varepsilon}{\varepsilon_1 A_1}\right) \leq \frac{1-P}{2} \tag{3.30}$$

in which Q stands for the normalized form of the cumulative normal distribution function, which can be solved to find the minimum required number of random experiments R .

So, following this approach, if one uses an appropriate number R of random phase realizations of the input multi-sine, the determination of all the coefficients is guaranteed, with P probability, to have an error smaller than ε.

An additional comment is imposed on (3.27). Some of the Ψ_{nl} were defined to produce zero output at the orthogonality level. Therefore, this expression can not be used directly to extract the coefficients relative to those terms. To extract those coefficients extra measurements must be performed with different input amplitudes so that the desired model terms produce a measurable effect.

After those measurements, an intermediate output signal is computed with the terms and coefficients already measured and this intermediate predicted signal is subtracted from the output. The difference between measured output and intermediate predicted output is then used to extract the non-orthogonal terms. Typical values of the number of randomizations required are shown in [3.7].

3.6 Passing from the Orthogonal Model to the Volterra Series

At this point the coefficients of a model able to mimic the system's response are known. However, the goal of this work is to determine the Volterra coefficients of (3.2). The procedure to compile those coefficients is conceptually simple. Since the orthogonal model was obtained recombining the Volterra series products, if the inverse combination is performed, then the original coefficients will be achieved. A simple way to accomplish this is to apply the probing method to (3.23) and calculate the model's response to each complex exponential excitation. Direct use of the probing method in (3.23) is not possible since the model basis functions $x_k(s)$ and $x_k'(s)$ are unable to represent complex signals. So

the basis functions $x_k(s)$ and $x_k^{'}(s)$ have to be redefined as:

$$x_k = A_k \cos(\omega_k t + \phi_k) + j.A_k \sin(\omega_k t + \phi_k), \tag{3.31a}$$

and the corresponding 90º deviation

$$x_k^{'} = A_k \sin(\omega_k t + \phi_k) - j.A_k \cos(\omega_k t + \phi_k) \tag{3.31b}$$

to allow the representation of the complex exponentials.

With these new basis functions in the Ψ's evaluation on (3.23) the mapping between the orthogonal model formulation and the Volterra series can be computed. After a lengthy and time consuming process of substitutions and simplifications the expressions for each of the frequency domain Volterra kernels are obtained.

Bellow the expressions for the Volterra series kernels determination from the orthogonal model representation are shown, for a generic number of input tones and a nonlinearity order up to 5.

$$H_0 = C_0 - A_0^2 \sum_{j=1}^{N} C_{I,j-j} + A_0^4 \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} C_{I,j_1-j_1 j_2-j_2} \tag{3.32a}$$

$$\begin{aligned} H_1(\omega_i) = {} & C_{I,i} - A_0^2 \sum_{i_1=1}^{N} C_{I,ii_1-i_1} + A_0^4 \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} C_{I,ij_1-j_1 j_2-j_2} \\ & - j.\left(C_{Q,i} - A_0^2 \sum_{i_1=1}^{N} C_{Q,ii_1-i_1} + A_0^4 \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} C_{Q,ij_1-j_1 j_2-j_2} \right) \end{aligned} \tag{3.32b}$$

$$H_2(\omega_{i_1}, \omega_{i_2}) = C_{I,i_1 i_2} - A_0^2 \sum_{j=1}^{N} C_{I,i_1 i_2 j-j} - j.\left(C_{Q,i_1 i_2} - A_0^2 \sum_{j=1}^{N} C_{Q,i_1 i_2 j-j} \right) \tag{3.32c}$$

$$H_3(\omega_{i_1}, \omega_{i_2}, \omega_{i_3}) = \frac{4}{6}\left[C_{I,i_1 i_2 i_3} - A_0^2 \sum_{j=1}^{N} C_{I,i_1 i_2 i_3 j-j} - j.\left(C_{Q,i_1 i_2 i_3} - A_0^2 \sum_{j=1}^{N} C_{Q,i_1 i_2 i_3 j-j} \right) \right] \tag{3.32d}$$

$$H_4(\omega_{i_1}, \omega_{i_2}, \omega_{i_3}, \omega_{i_4}) = \frac{8}{24}\left(C_{I,i_1 i_2 i_3 i_4} - j.C_{Q,i_1 i_2 i_3 i_4} \right) \tag{3.32e}$$

$$H_5(\omega_{i_1}, \omega_{i_2}, \omega_{i_3}, \omega_{i_4}, \omega_{i_5}) = \frac{16}{120}\left(C_{I,i_1i_2i_3i_4i_5} - j.C_{Q,i_1i_2i_3i_4i_5}\right) \quad (3.32f)$$

The Volterra kernels presented in these expressions can be converted into their equivalent time domain representation using conventional multidimensional Fourier Transforms. After all this process the desired coefficients of equation (3.2) are obtained.

In expressions (3.32), it can be seen that the determination of the first order Volterra kernel is dependent on the higher order orthogonal kernels. This dependency is already known [3.1], and arises from the fact that the orthogonal first order coefficients are not the small signal 1st degree term, but actually, a Best Linear Approximator of the system (thus containing components of higher orders, as discussed on Chapter 5). Since the Volterra series is homogeneous, the higher order components included in the first order terms must be subtracted when writing the Volterra series coefficients.

3.7 Summary and Conclusions

In this chapter the whole procedure to reach the orthogonal extraction of the Volterra series coefficients was presented. The formal equations to build such an orthogonal model for an input randomized phase multisine were derived and explained and an expression for the general N^{th} order term was given.

The orthogonal coefficients' extraction was explained and closed form expressions derived to determine the minimum number of different input randomized phase multisine realizations that ensures a desired maximum error for coefficients' determination.

Finally, the correspondence between the orthogonal model coefficients and the conventional Volterra series ones was derived.

The next chapter will present several application examples of this coefficient extraction methodology.

3.8 References

[3.1] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: John Wiley & Sons, 1980.

[3.2] S. Boyd and L. Chua, "Fading Memory and the Problem of Approximating Nonlinear Operators with Volterra Series," *IEEE Trans. on Circuits & Systems*, vol. CAS-32, pp. pp.1150-1161, 1985.

[3.3] V. Mathews and G. Sicuranza, *Polynomial Signal Processing*. New York: John Wiley & Sons, Inc, 2000.

[3.4] R. Pintelon and J. Schoukens, *System Identification - A Frequency Domain Approach*: IEEE Press, 2001.

[3.5] P. M. Lavrador, N. B. Carvalho, and J. C. Pedro, "Evaluation of signal-to-noise and distortion ratio degradation in nonlinear systems," *IEEE Trans. on Microwave Theory and Techniques*, vol. MTT - 52, pp. 813-822, 2004.

[3.6] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley Publishing Company, 1994.

[3.7] P. M. Lavrador, J. C. Pedro and N. B. Carvalho, "A new method for the orthogonal extraction of the Volterra series coefficients," *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits Proceedings CDROM,* Aveiro, Portugal, Jan. 2006.

4. Approximation results with the new model formulation in different conditions

4.1 Introduction

In this chapter a demonstration of the performance of the orthogonal modelling approach presented in Chapter 3 is shown. To accomplish this demonstration in a complete way, situations will be shown where the model approximation is very good, but also some other examples in which this modelling approach is not recommended.

To circumvent practical limitations on the number of extracted coefficients and of signal time samples required to represent a real communications' system, a lowpass equivalent model formulation will be used. This way the number of time samples required decreases a lot since the sample frequency is chosen to verify the Nyquist criteria for the signal's bandwidth and not for the carrier frequency.

With the purpose of demonstrating the model's ability to deal with the different phenomena present in a real microwave PA, the modelling approach presented in the

previous chapter will here be exemplified and demonstrated for different kinds of systems [4.1].

This chapter starts with two control tests: (i) extracting the model of a nonlinear memoryless amplifier modelled by a mathematical function (ii) and extracting the model of a linear band pass filter. The purpose of these simple situations is to illustrate the modelling procedure and to create control experiences that will allow the verification of the model performance when modelling more realistic systems. After these basic initial control tests some linear memory was cascaded with a nonlinear memoryless system creating, this way, the Wiener, Hammerstein and Wiener-Hammerstein configurations. After these tests, an extra synthetic amplifier was used: a nonlinear amplifier with memory effects caused by the bias circuitry. The device used to test this configuration was a real power amplifier modelled by its equivalent ADS [4.2] representation, and the model results obtained by the proposed modelling approach are compared with the ones obtained in ADS.

Finally, real amplifiers are measured and a model is extracted for two different amplifiers, one of them presenting strong nonlinear memory effects due to a badly designed (on purpose) bias circuits.

For all these different power amplifier situations some figures are presented in order to characterize their operating point and to demonstrate the model's performance. Also, for each case the one tone input/output power curve is shown to characterize the nonlinearity of the amplifier used. To illustrate the linear memory present in the system the small signal transfer function is plotted. Finally, to characterize the nonlinear memory, the instantaneous input/output power plot is also shown.

4.2 First Example: A Memoryless Amplifier

Figure 4.1 presents the block diagram of the nonlinear system being modelled in this example. A nonlinear function (in this case an hyperbolic tangent) converts the input time signal $x(t)$ into its correspondent output signal $y(t)$.

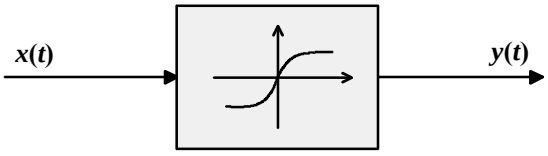


Figure 4.1 – Block diagram of the nonlinear memoryless system considered.

The particular expression implemented in this example was:

$$y(t) = -10 \tanh\left(\frac{x(t)}{1.3}\right) \tag{4.1}$$

in which $x(t)$ and $y(t)$ are the input and output voltages respectively.

As was stated above, the goal of modelling this memoryless amplifier is to illustrate the steps required to extract the Volterra series model. So, to find the required model parameters (order of the polynomial and the number of input tones of the multisine) we must start with some raw measurements in order to observe, qualitatively, the behaviour of the system.

Figure 4.2 shows the system's small signal transfer function. This figure allows the evaluation of the linear memory present in the system. In this simple case, since a memoryless system was created on purpose, the transfer function is flat over frequency.

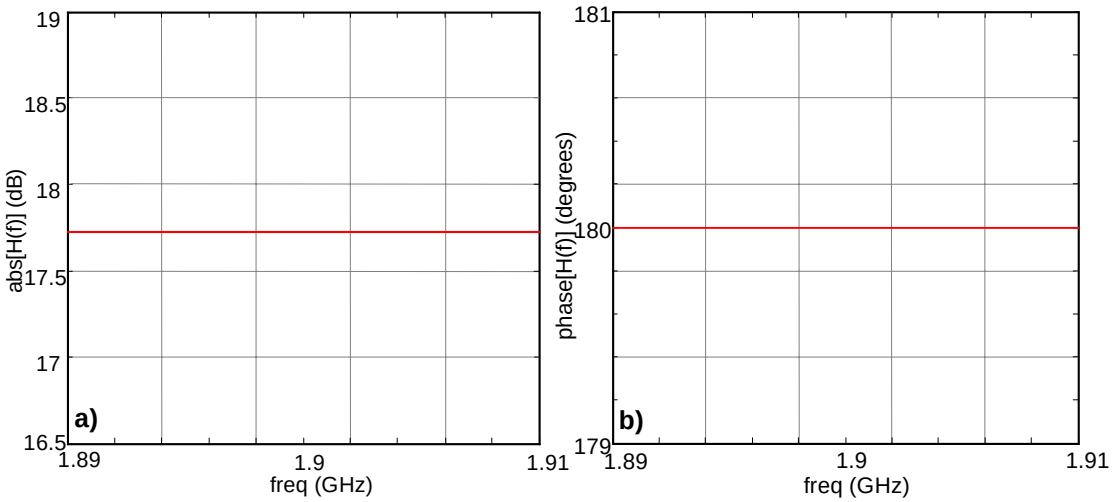


Figure 4.2 – Small signal frequency response of the memoryless amplifier. a) amplitude b) phase.

Figure 4.3a) illustrates the one tone input/output average power. As a reference, it is also shown the response of a linear system with identical small signal gain. One extra step is required to complete our initial system analysis – it is the instantaneous input/output power gain plot. This test shows the dynamic behaviour of the systems. The presence of hysteresis [4.3] on this curve might indicate linear memory (if it appears for low input signal power) and/or nonlinear memory (if it appears for high input signal power). In this case, histeresis does not appear either in small or large signal regime, as was expected for this memoryless system.

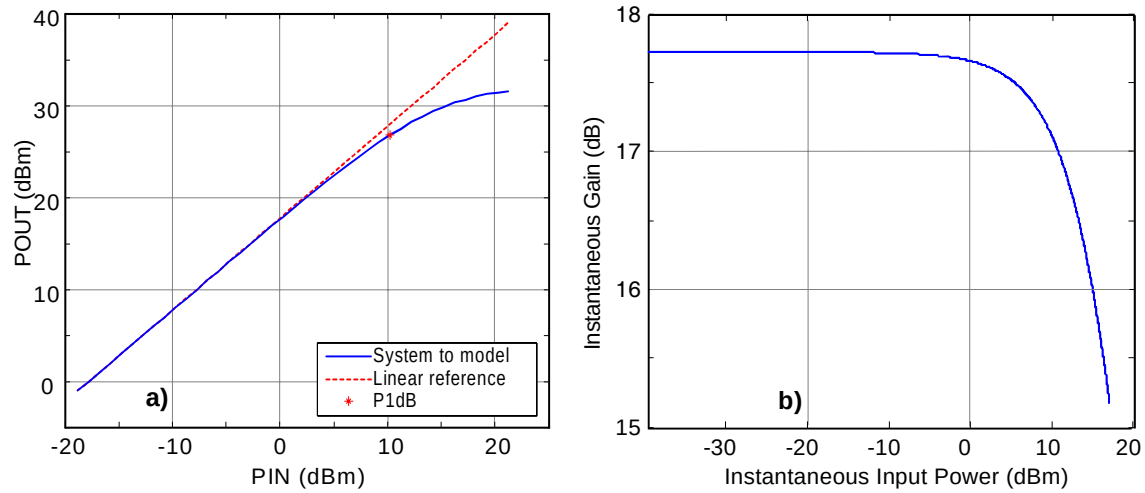


Figure 4.3 – a) One tone input/output average power. b) Dynamic gain curve of the amplifier considered (obtained with two tones)

After these preliminary tests, the number of input tones required for model extraction can be determined. Since the system in this example presents no memory effects, a single tone measurement is sufficient for model extraction.

Figure 4.3 shows that the input one dB compression point of the system is approximately 10 dBm. The model was extracted with an average input power of 8 dBm.

Figure 4.4 shows a comparison between observed and predicted input/output one tone power curves. Very good results are seen up to an average input power of 18dBm. Above this point the polynomial divergence phenomena starts to appear. It must be pointed out that we are modelling a hyperbolic tangent system using a fifth order polynomial. So, sooner or later the model should diverge. However, when the model diverges, the input power level is more than 6 dB above the extraction point, and the system is already in deep compression (more than 3 dB).

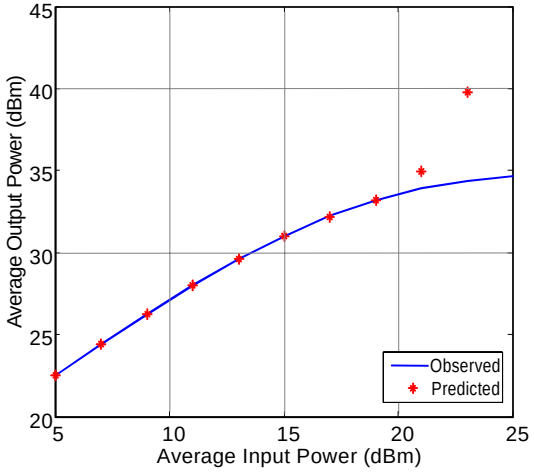


Figure 4.4 – Input/output average power for one tone signal: observed and modelled results.

Additionally to the one tone input/output validations, another test was performed that consisted in evaluating the model's accuracy when excited with a WCDMA signal. As an indicative example, it is shown in Figure 4.5a) and Figure 4.5b) the time domain complex envelope and spectrum, respectively, of the WCDMA sequence used. The average input power of the signal used was equal to the model extraction point.

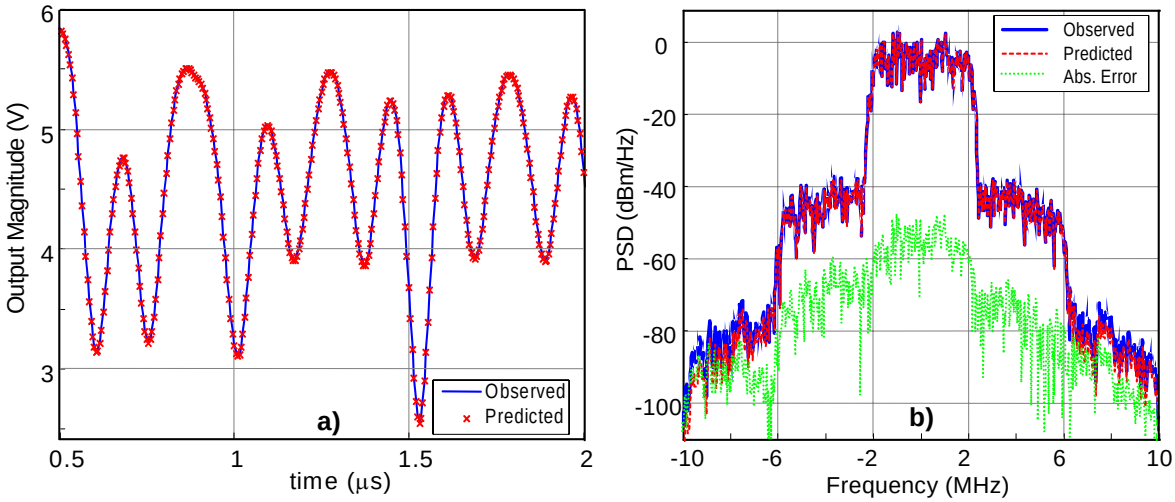


Figure 4.5 – a) Output magnitude of the complex envelope of the WCDMA sequence used to validate the model. b) – Observed and predicted power spectral density of the WCDMA signal used to validate the model.

In Figure 4.5b), the WCDMA power spectrum is shown. In-band results are nearly coincident. The absolute error between both is of the order -50 dBm. On the contrary, the alternate channel approximation is not as good and at some points the value of the error observed is of the same order of magnitude as the modelled signal. This result is probably originated on the power level difference from the in-band component (nearly -10dBm) and the fifth order one (around -80dBm). This power level difference complicates the accurate modelling of the small power IMD components.

Another validation test performed is the comparison shown in Figure 4.6 that plots the dynamic input/output characteristics measured and modelled. In this example the modelled curve presents a wider region in the small input signal power. This can be interpreted as the impact of the modelling noise (which is more visible when the signal contribution is negligible).

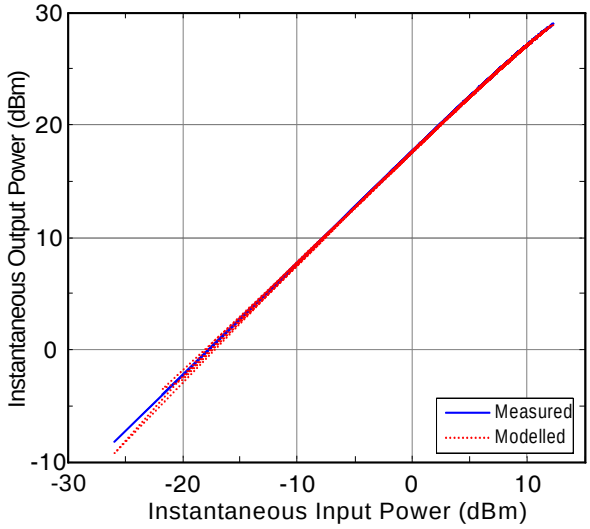


Figure 4.6 – Observed and Predicted instantaneous input/output transfer characteristic for the memoryless system.

Additionally, the comparison of the output in-band, adjacent and alternate channel power modelled and measured is shown for a WCDMA validation signal. The plot in Figure 4.7 shows an overview of all this data and a good agreement is seen. In an ideal situation, the alternate channel power (IMD5) would raise at 5dB/dB. It is not visible in this figure because the low power fifth order IMD is masked by the spectral leakage of the signal.

Finally, the Normalized Mean Square Error (NMSE) is plotted in Figure 4.7b). A NMSE smaller than -40 dB can be seen for power levels bellow the extraction level. The model divergence for high input power levels causes the fast degradation of the NMSE.

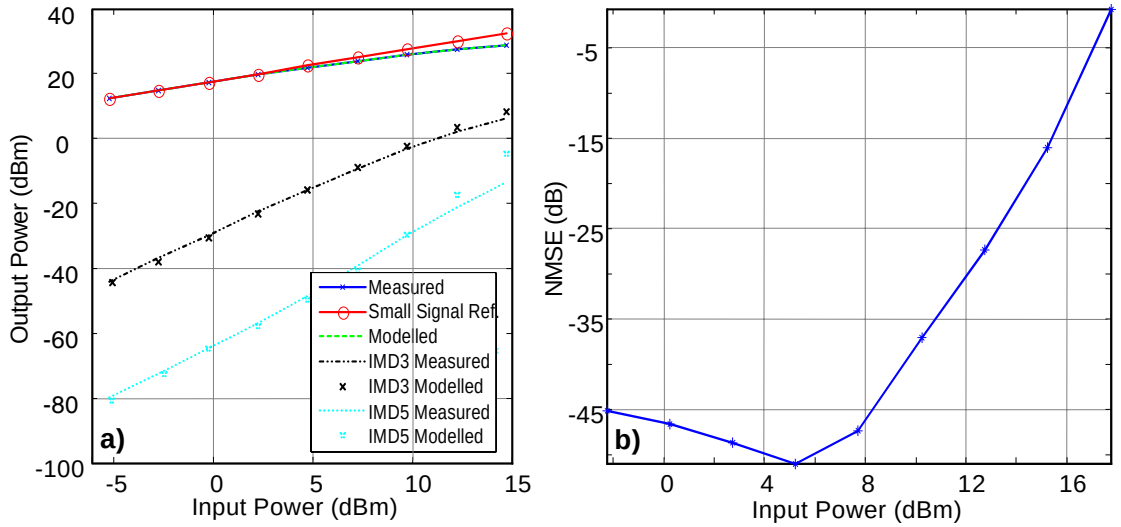


Figure 4.7 – a) Comparison between measured and modelled output power, IMD3 and IMD5. b) Normalized Mean Square Error variation with the input power sweep.

4.3 Modelling a Linear Filter

The schematic representation of the system used in this modelling example is presented in Figure 4.8. The adopted filter was a 4th order Butterworth bandpass. The frequency response of this filter is shown in Figure 4.9.

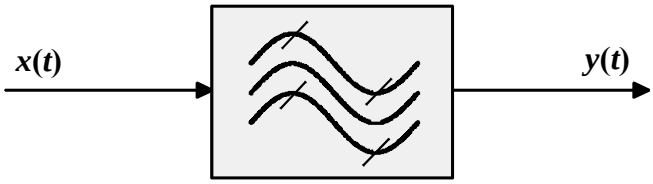


Figure 4.8 – Schematic representation of the linear filter used.

The linear filter transfer function is presented in the following equation

$$H(z) = 10^{-5} \frac{0.3969 - 0.7938z^{-2} + 0.3969z^{-4}}{1 - 3.3054z^{-1} + 4.7278z^{-2} - 3.2995z^{-3} + 0.9964z^{-4}} \qquad (4.2)$$

Since, in this situation, a linear system is to be modelled the nonlinear system characterization figures will not be explored. So, the preliminary characterization of the linear filter is made by its small signal frequency response presented in Figure 4.9. Despite the constant amplitude response in a bandwidth of nearly 5 MHz, it is visible in the phase variation the non-constant frequency response of the filter inside that band.

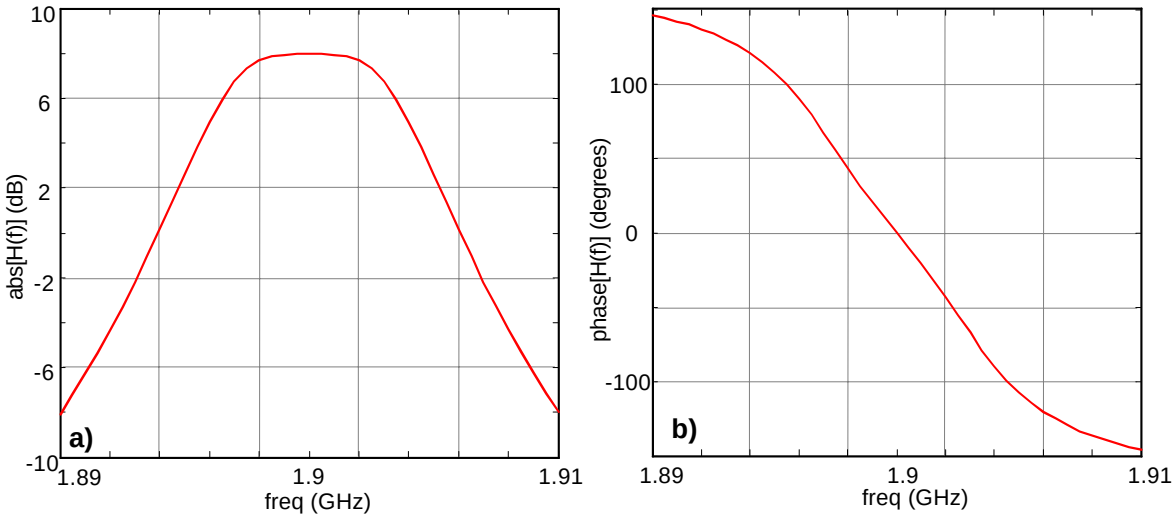


Figure 4.9 – Frequency response of the linear filter used in this situation.

Figure 4.10a) shows the one tone input/output average power in which the system's linearity can be seen by the superposition of the measured line with the reference linear system.

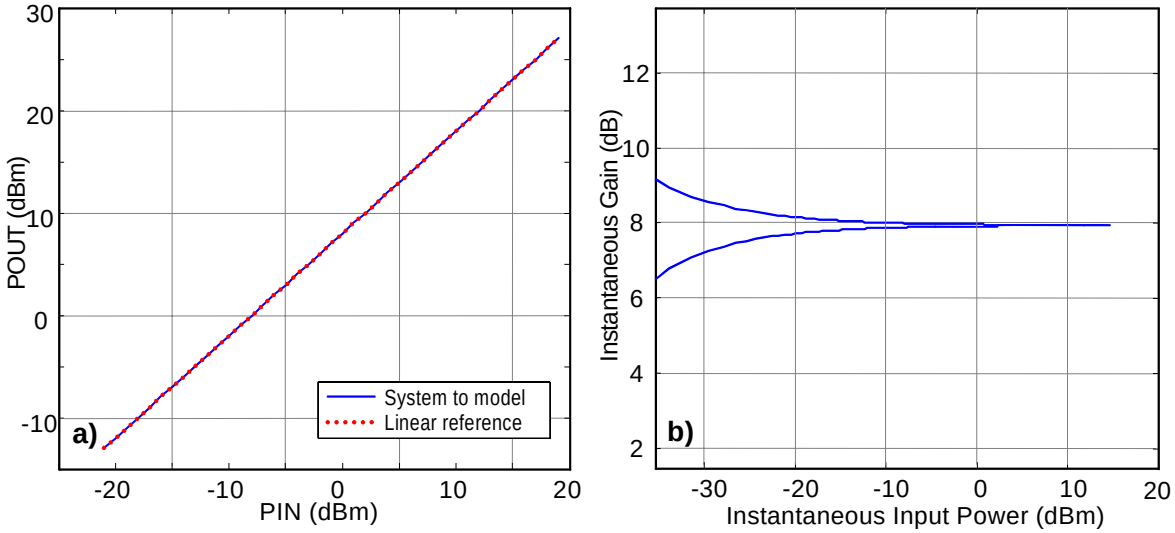


Figure 4.10 – a) One tone input/output average power of the linear system. b) Dynamic gain curve of the linear filter.

Figure 4.10b) shows the instantaneous gain of the filter used. It is shown that the system presents a linear behaviour. The aperture of the curve in the left side of the figure (for very low input powers) demonstrates some delay on the zero crossing zones imposed by the linear memory of the filter.

In order to determine the number of samples (frequency tones) required to model this filter an estimation of its lowpass equivalent impulse response is made using a large number of frequency tones (15), the result is shown in Figure 4.11.

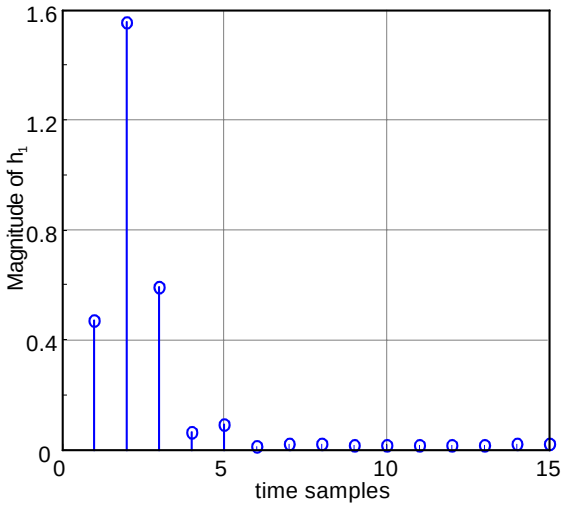


Figure 4.11 – The magnitude plot of the linear filter impulse response.

From a rough analysis of this figure one can realise that the 15 time samples are more than enough to contain the more significant part of the filter's impulse response energy. Since the most significant part of the energy is concentrated on the first five samples this is the number considered for model extraction.

So the nonlinear model is extracted with five tones up to fifth order, and its modelling performance is presented in the next figures. In Figure 4.12a) the comparison between modelled and observed input/output is shown for a single tone. Since the system under test is, in this case, linear, the model does not present deviations when the input power leaves the model extraction level.

Figure 4.12b) shows that there is a good convergence between measured and modelled instantaneous output waveforms.

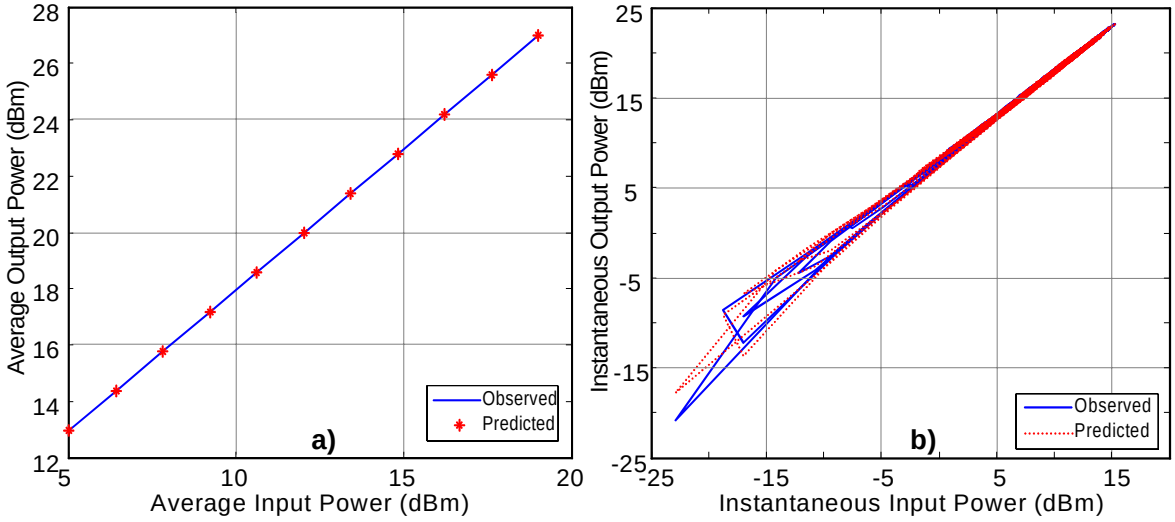


Figure 4.12 – a) Input/output power comparison between observed and predicted waveforms for an input multisine. b) Predicted and observed instantaneous input/output transfer characteristic for the linear filter. (obtained with a WCDMA signal)

As validation, in Figure 4.13 the modelling results of a WCDMA signal are plotted in time and frequency domain, respectively.

In Figure 4.13b), differences are shown for the adjacent and alternate channels. However the spectrum in those regions is more than 100 dB bellow the in-band power and thus the observed differences do not have physical meaning.

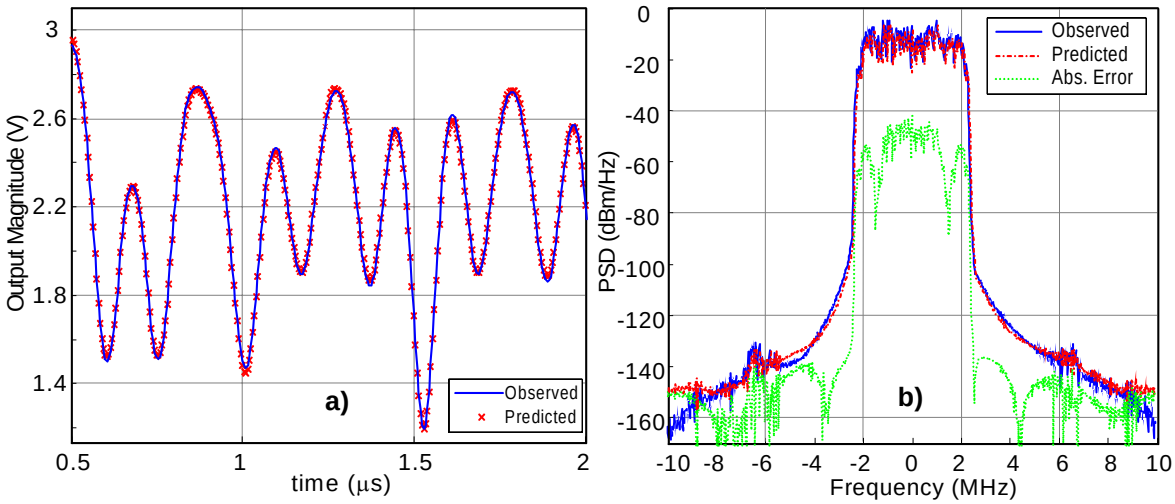


Figure 4.13 – a) Comparison between observed and predicted magnitude of the complex envelope of the WCDMA sequence used to validate the model. b) The predicted and observed spectra of WCDMA validation signal.

Figure 4.14a) shows the variation of in-band, adjacent and alternate channel powers both measured and modelled with an input WCDMA signal power sweep.

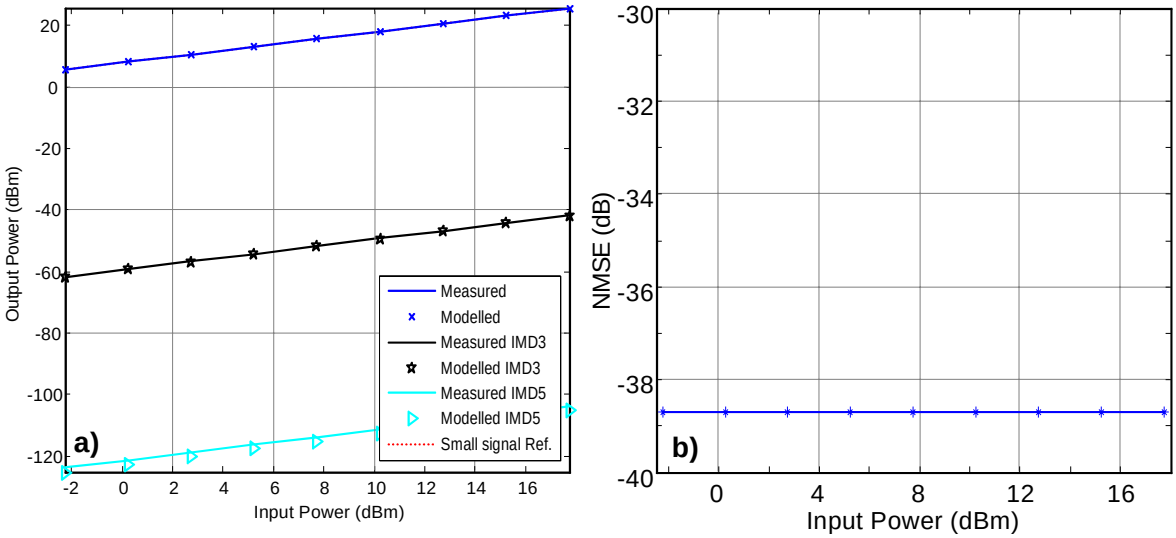


Figure 4.14 – a) Comparison between measured and modelled output power, IMD3 and IMD5 for the linear filter system. b) – Normalized Mean Square Error evolution with input power sweep.

Since in this particular example a linear system is being modelled, the nonlinear coefficients obtained are almost zero (the differences are the extraction noise), there is no NMSE degradation with the input power variation, as shown in Figure 4.14b.

4.4 Modelling the Cascade of a Memoryless Nonlinearity and a Linear Filter

4.4.1 Wiener Configuration

In this sub-section an amplifier in a Wiener configuration is considered. Figure 4.15 shows the nonlinearity following the linear filter. It can be understood as the sequence of two systems of the type of the ones that were previously studied.

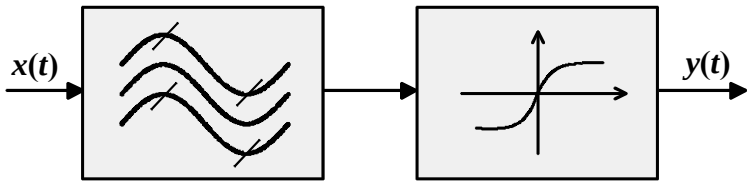


Figure 4.15 – Schematic representation of the Wiener model amplifier configuration.

For this system the small-signal S_{21} parameters are presented in Figure 4.16, in both, amplitude and phase.

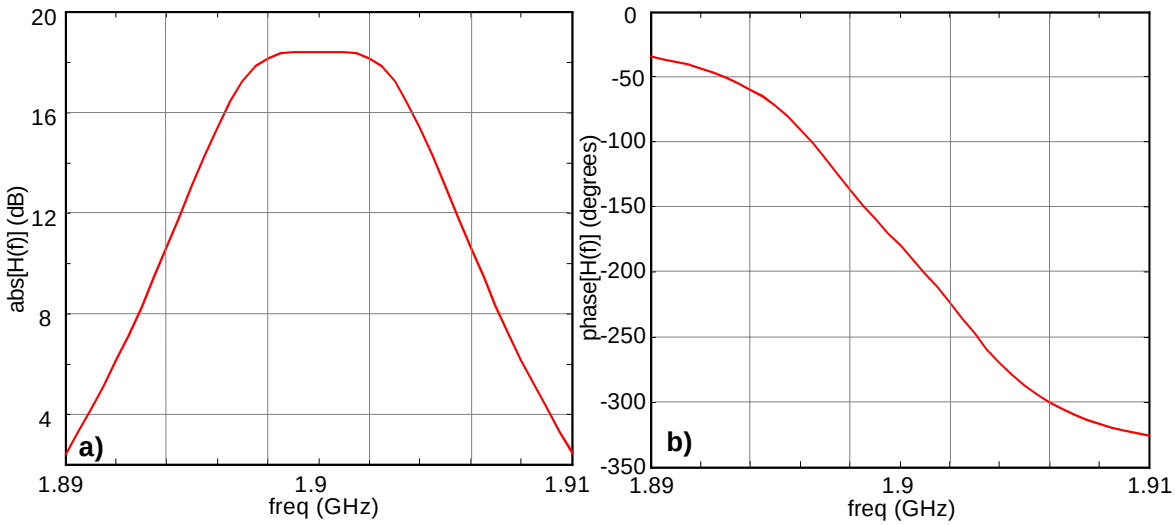


Figure 4.16 –Small signal gain variation with frequency of the Wiener system considered. a) amplitude; b) phase.

The one tone input/output power is plotted in Figure 4.17a), the one dB compression point is also indicated in the figure.

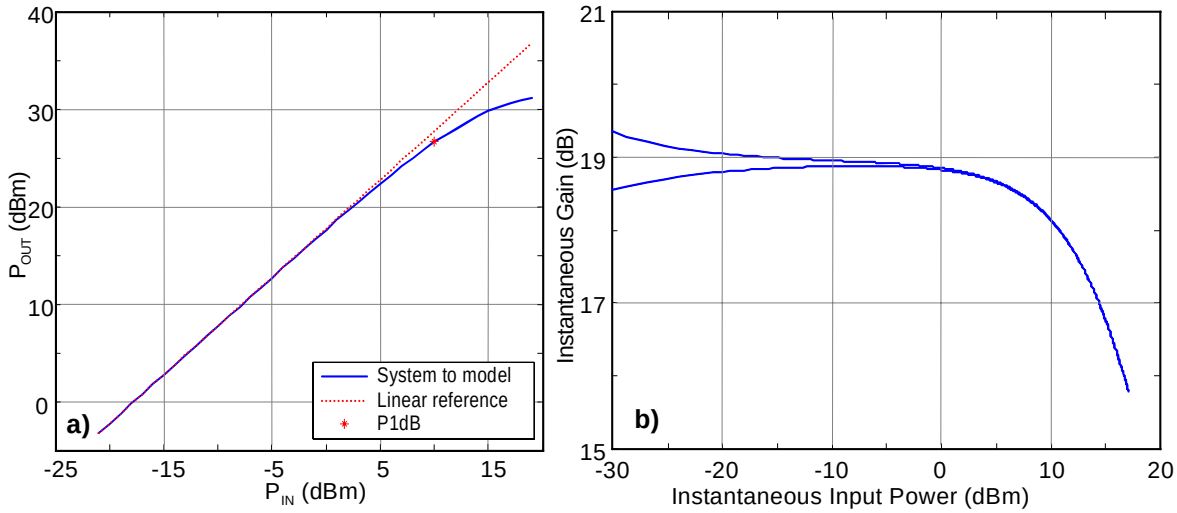


Figure 4.17 – a) One tone average P_{IN}/P_{OUT} of the system of Figure 4.15. b) Instantaneous two-tone gain curve for the Wiener configuration.

Figure 4.17b) shows the instantaneous two-tone gain characteristic of the Wiener system under modelling in this section. The right part of the figure shows the nonlinear behaviour of the system, while the left part shows the memory behaviour visible for the small signal regime - thus is linear memory.

The model for this system was obtained with five time delays and fifth order. Figure 4.18a) shows the comparison between modelled and measured single tone output power. In this situation the polynomial divergence for high input powers is also shown. Figure 4.18b) shows a good convergence of the instantaneous input/output power shadows for both modelled and measured signals.

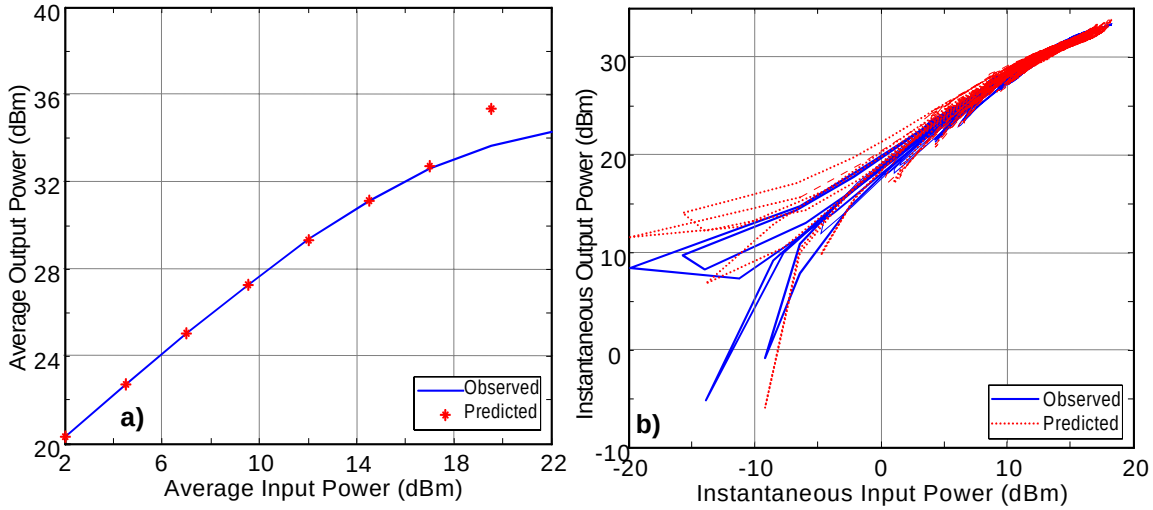


Figure 4.18 – a) Single tone input/output average power (modelled and observed). b) Comparison between modelled and measured WCDMA instantaneous output for the Wiener system.

Figure 4.19a) shows a comparison between modelled and measured WCDMA signal complex envelope magnitude. It is seen that in this case the model has some difficulties to track the measured output. Nevertheless, Figure 4.19b), which shows the same results in the frequency domain, presents an absolute error of more than 20 dB bellow the output power level for the in-band channel, and a similar result for the adjacent channel (dominated by third order components). The fifth order components present an absolute error closer to the actual value of the IMD components. That might be due to the low level of these components.

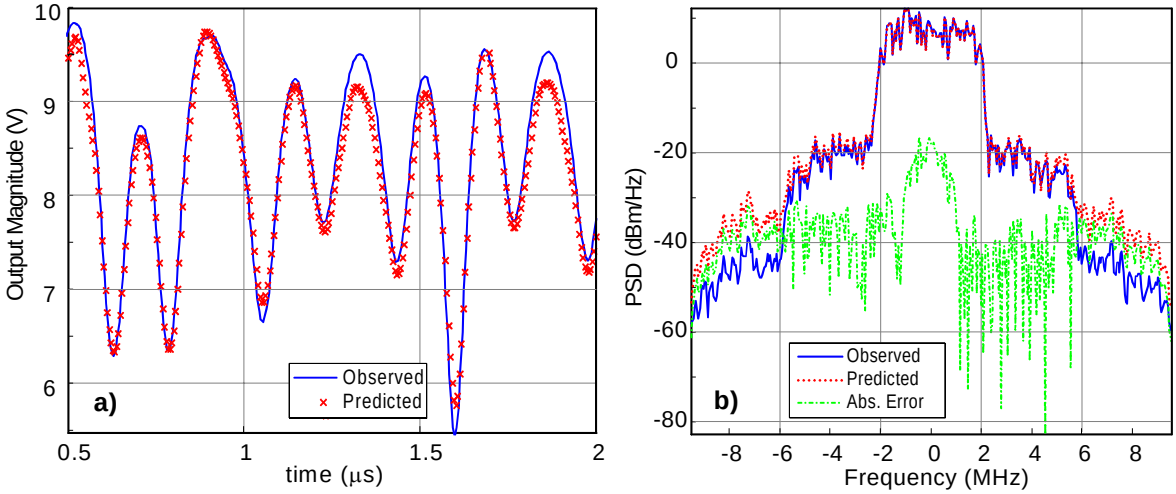


Figure 4.19 – Magnitude of the time domain complex envelope of the WCDMA signal (Measured and Modelled). b) – The predicted and observed spectra of CDMA validation signal.

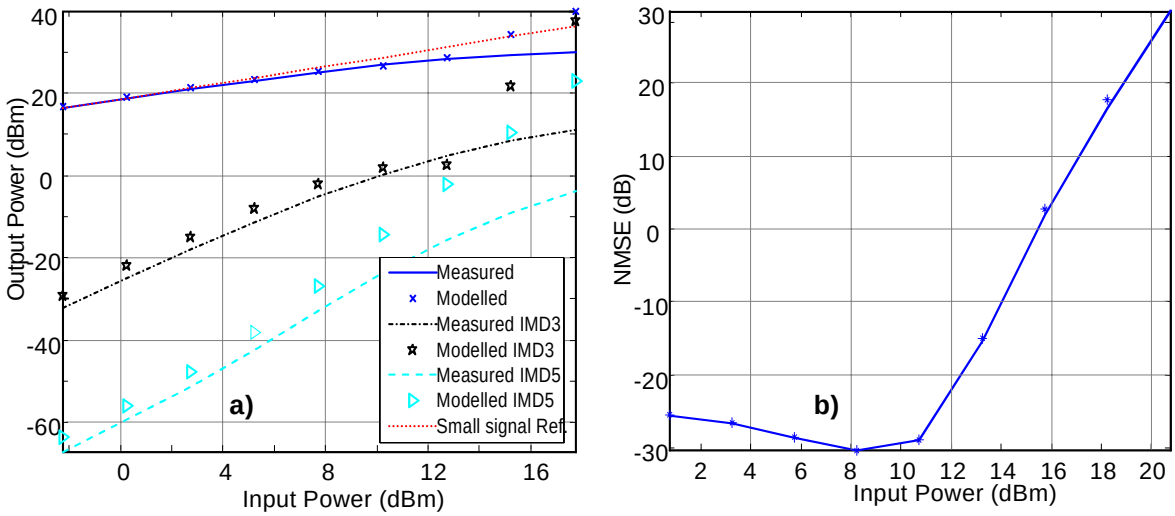


Figure 4.20 – a) Model results on output power, IMD3 and IMD5 with input WCDMA signal power sweep. b) Normalized Mean Square Error of the model of the Wiener system.

Figure 4.20a) presents the effect of input power sweep on the in-band, adjacent and alternate channel power predictions. Once more it is shown that above the extraction level the model rapidly diverges from the measured values. In this figure this is particularly evident for the third order components.

In this situation the best NMSE achieved is -30 dB, as exposed in Figure 4.20b), which is not a very good value, but is in accordance with the imperfect approximation shown in Figure 4.19a).

4.4.2 Hammerstein Configuration

In this validation, a different test has been performed. The memoryless nonlinearity was placed before the nonlinear filter according to the Hammerstein configuration shown in Figure 4.21.

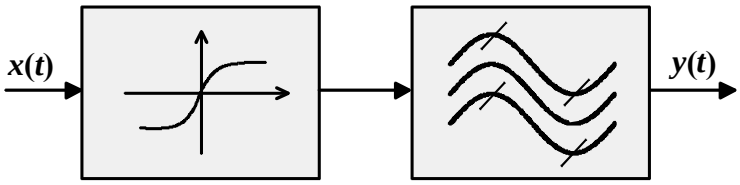


Figure 4.21 – Schematic representation of the Hammerstein model configuration.

For this system the small-signal S_{21} parameters are plotted in the next figure.

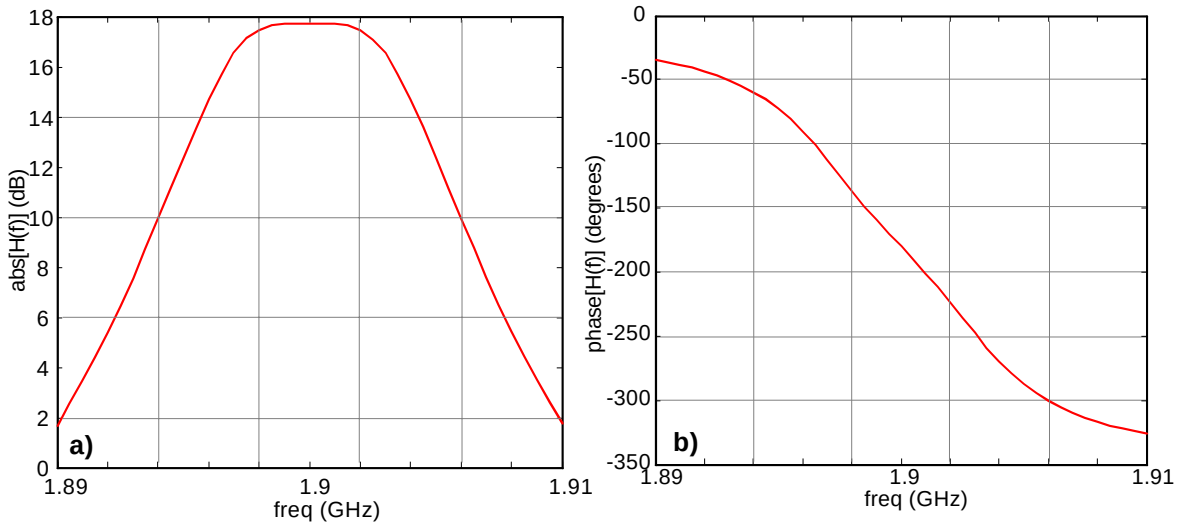


Figure 4.22 –Small signal gain variation with frequency of the Hammerstein system considered. a) amplitude; b) phase.

The one tone input/output system characteristic is exposed in Figure 4.23a). Also shown is the one dB compression point (where the model was extracted).

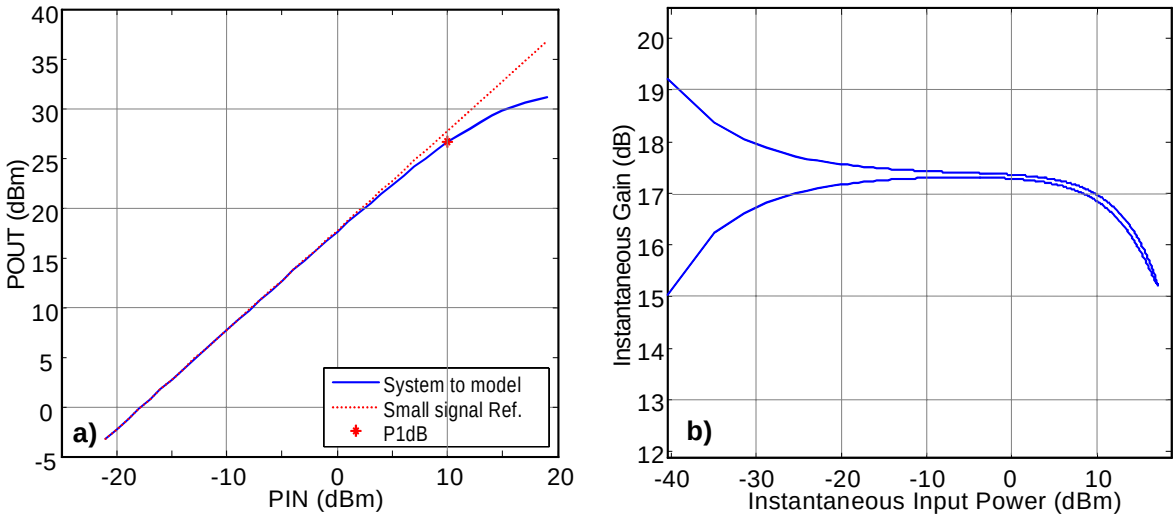


Figure 4.23 – a) One tone average P_{IN}/P_{OUT} of the system of Figure 4.21. b) Instantaneous two-tone gain curve for the Hammerstein configuration.

Figure 4.23b) shows the instantaneous two tone gain of this system. It can be seen that the system behaviour is mainly dominated by linear memory (low input power) and nonlinearity.

The system with the characteristics presented in the previous figures has been modelled and now the modelling performance will be evaluated. To start, Figure 4.24a) shows the comparison between measured and modelled one tone input/output power.

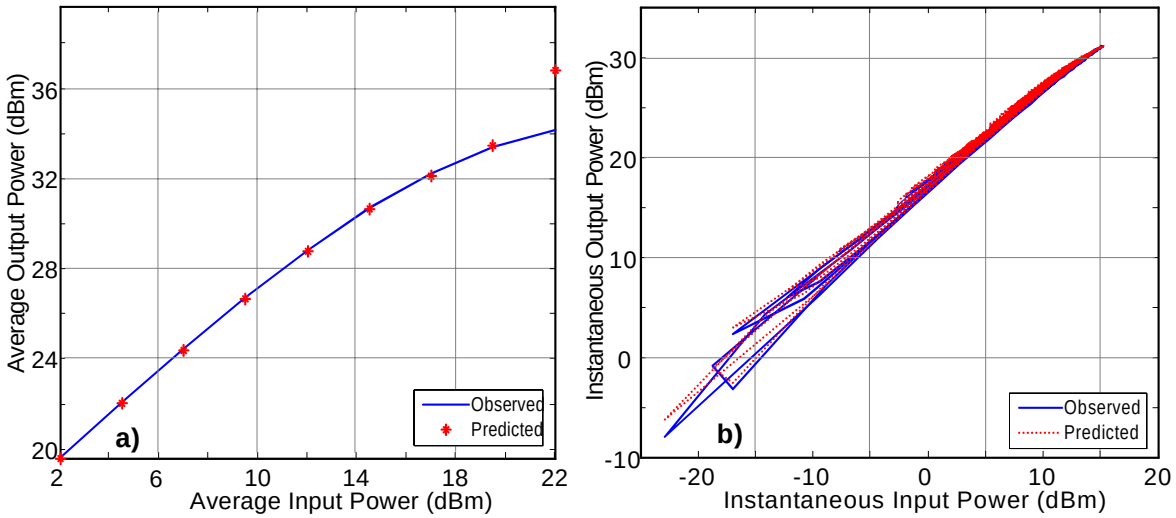


Figure 4.24 – a) Observed and Predicted one tone input/output average power. b) Comparison between observed and predicted WCDMA instantaneous output for the Hammestein system.

The cross validation results, with WCDMA sequence, are plotted in Figure 4.25a). It is shown that the model approximates the shape of the envelope amplitude but presents some difficulties to track it with detail.

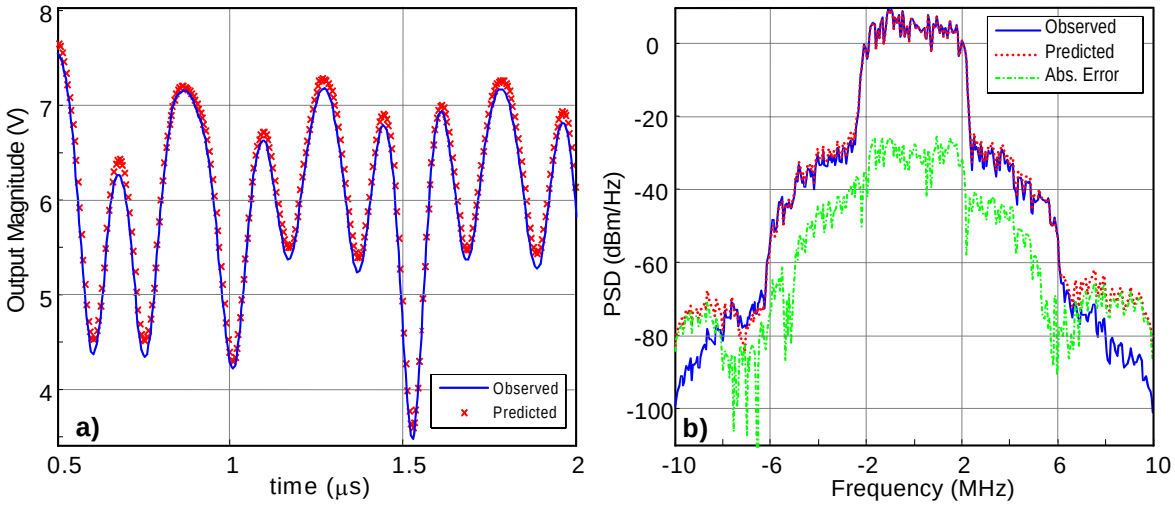


Figure 4.25 – a) Time domain complex envelope magnitude comparison between Measured and Modelled CDMA signals. b) The predicted and observed spectra of WCDMA validation signal.

Figure 4.25b) shows the correspondent frequency domain plot of the WCDMA envelope signal.

The next figure shows the variation of the in-band, adjacent and alternate channel integrated power both measured and simulated. As in the previous examples good agreement is shown except in the region of the polynomial model divergence.

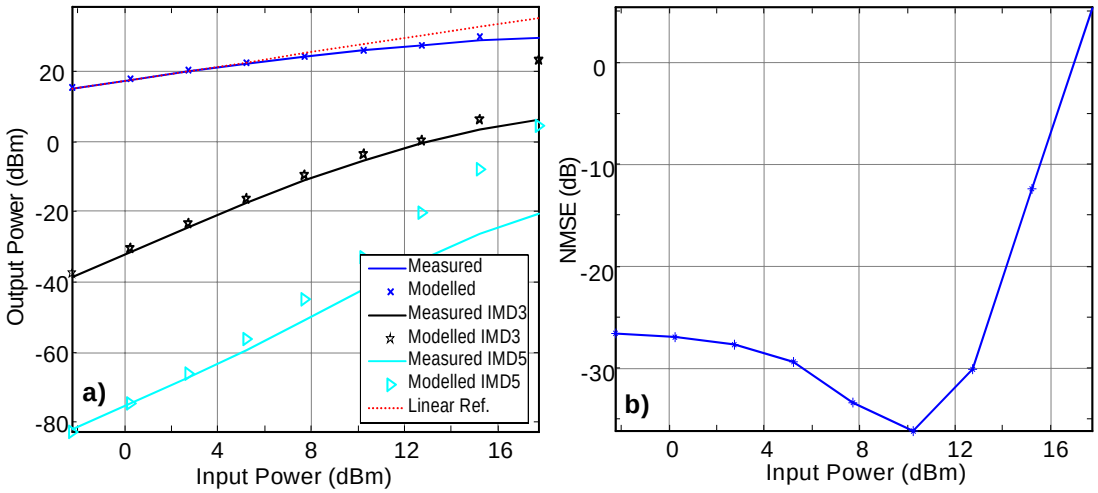


Figure 4.26 – a) Model results on output power, IMD3 and IMD5 with input WCDMA signal power sweep for the Hammerstein system. b) Normalized Mean Square Error of the Hammerstein system model results.

The NMSE plot of this model, Figure 4.26b), shows a clear valley that illustrates perfectly the local characteristic of the model being used in this thesis.

4.4.3 Wiener-Hammerstein Configuration

The block diagram of the Wiener Hammerstein configuration is shown in Figure 4.27.

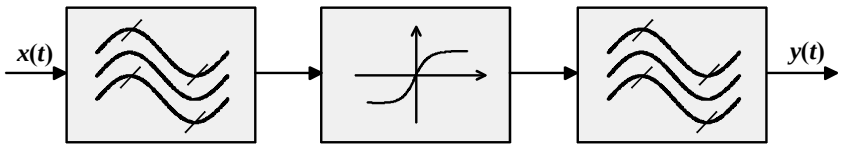


Figure 4.27 – Schematic representation of the Wiener-Hammerstein model configuration.

The small-signal S_{21} parameters, of this system are presented in Figure 4.28.

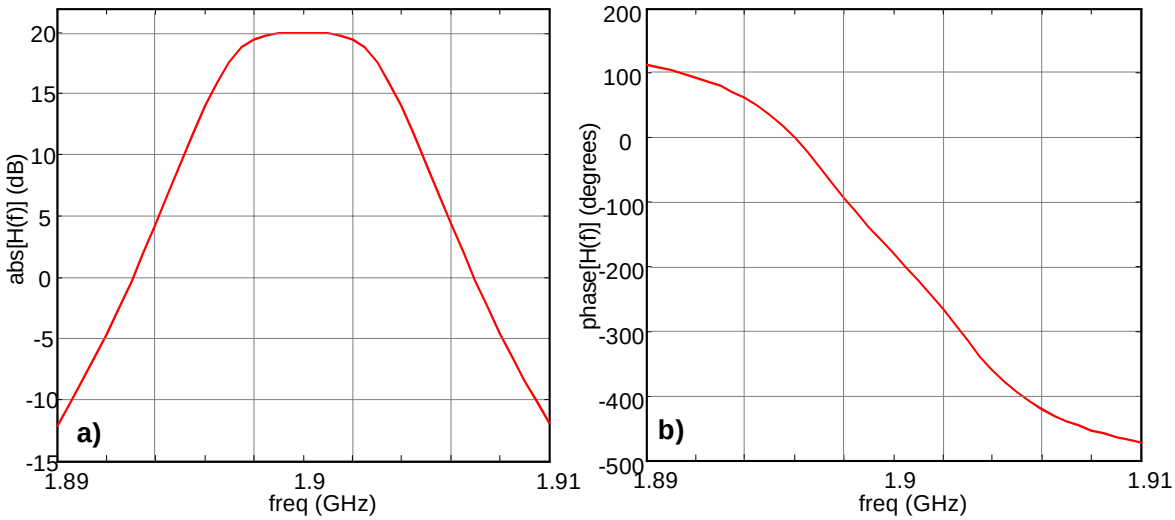


Figure 4.28 –Small signal gain variation with frequency of the Wiener-Hammerstein system considered. a) amplitude; b) phase.

The next figure shows the P_{IN}/P_{OUT} transfer characteristic of this system measured using a single tone. The dynamic transfer characteristic is plotted in Figure 4.29b).

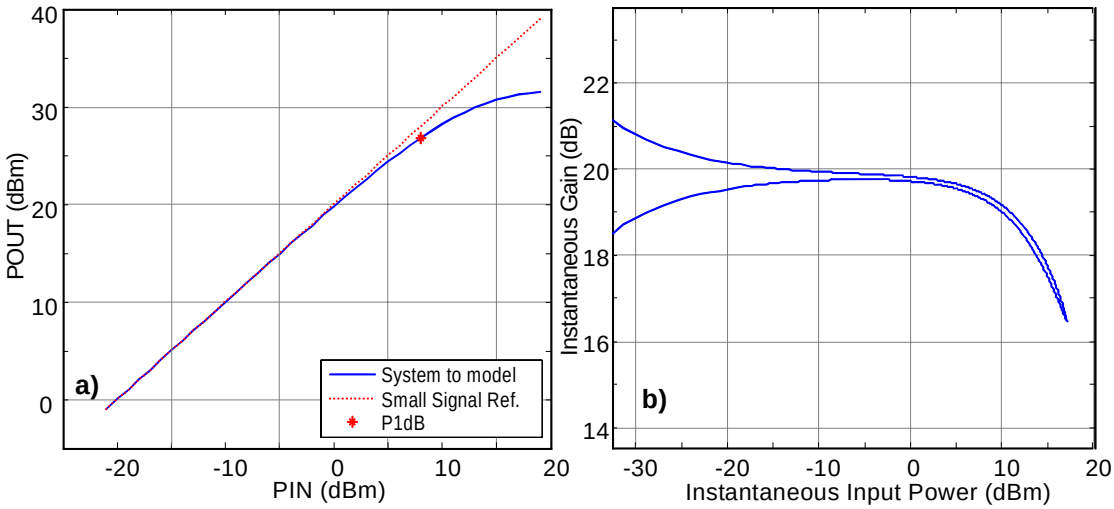


Figure 4.29 – a) One tone average P_{IN}/P_{OUT} of the system of Figure 4.27. b) Instantaneous two-tone Gain curve for the Wiener-Hammerstein configuration.

Figure 4.30a) presents the comparison of input/output power obtained in a static way using a single-tone excitation. Figure 4.30b) presents the measured/modelled instantaneous input/ouput power curves obtained with a WCDMA signal.

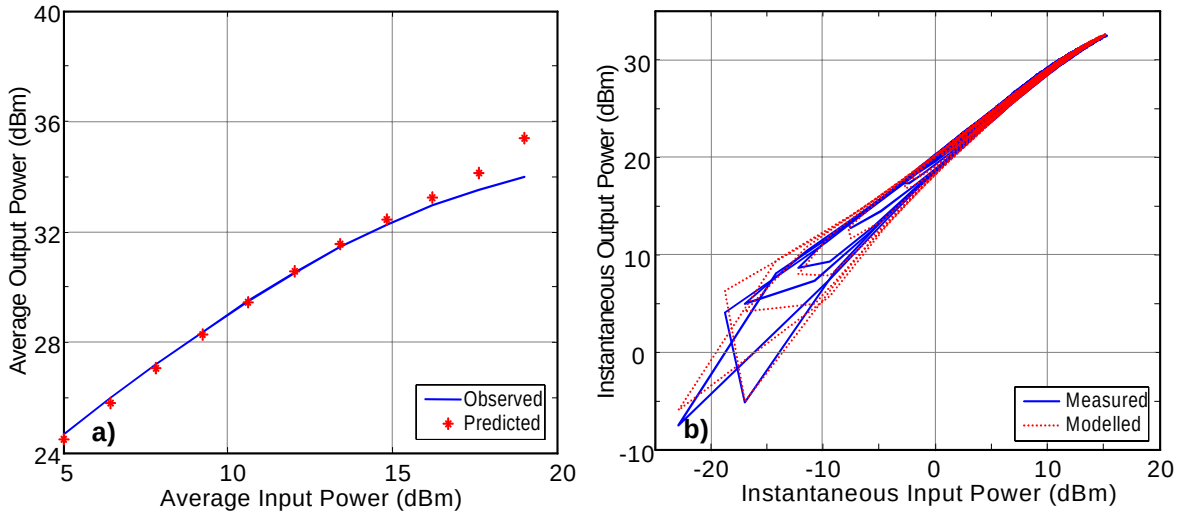


Figure 4.30 – a) Observed and predicted one tone input/output average power. b) Comparison between Modelled and Measured WCDMA instantaneous output for the Wiener-Hammerstein system.

Figure 4.31a) and Figure 4.31b) show the measured and modelled complex envelope of the WCDMA signal used to perform the cross validation of the model.

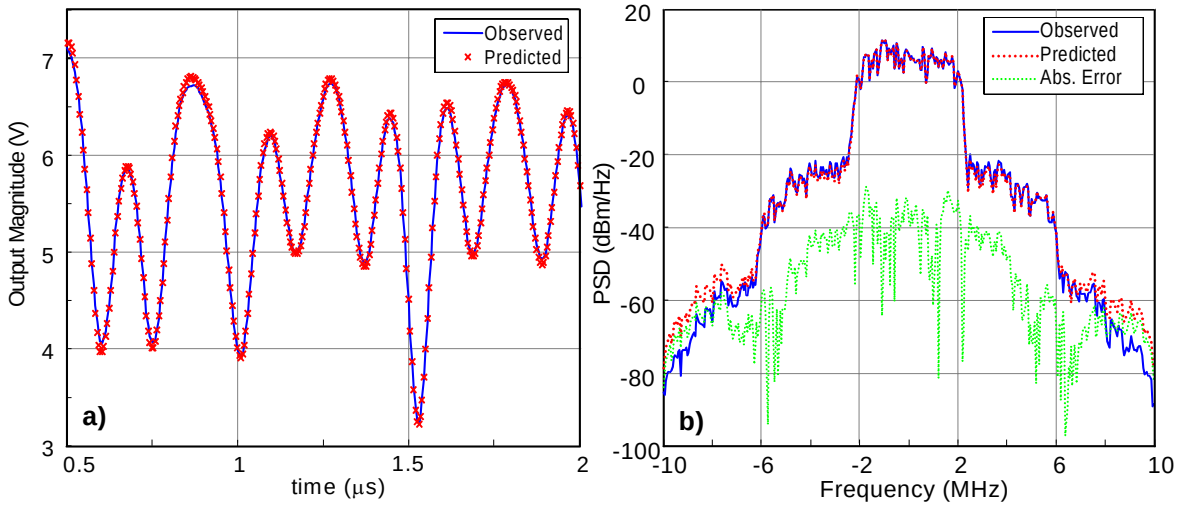


Figure 4.31 – a) Magnitude of complex time domain envelope comparison between Measured and Modelled CDMA signals. b) The predicted and observed spectra of CDMA validation signal.

Figure 4.32a) presents the evolution of the WCDMA integrated power with an input power sweep. Once again, the model divergence is quite clear for input power above the model extraction point.

Figure 4.32b) presents the variation of the NMSE for the Wiener-Hammerstein system model with the input power level variation. The model's local behaviour is once again visible on this figure.

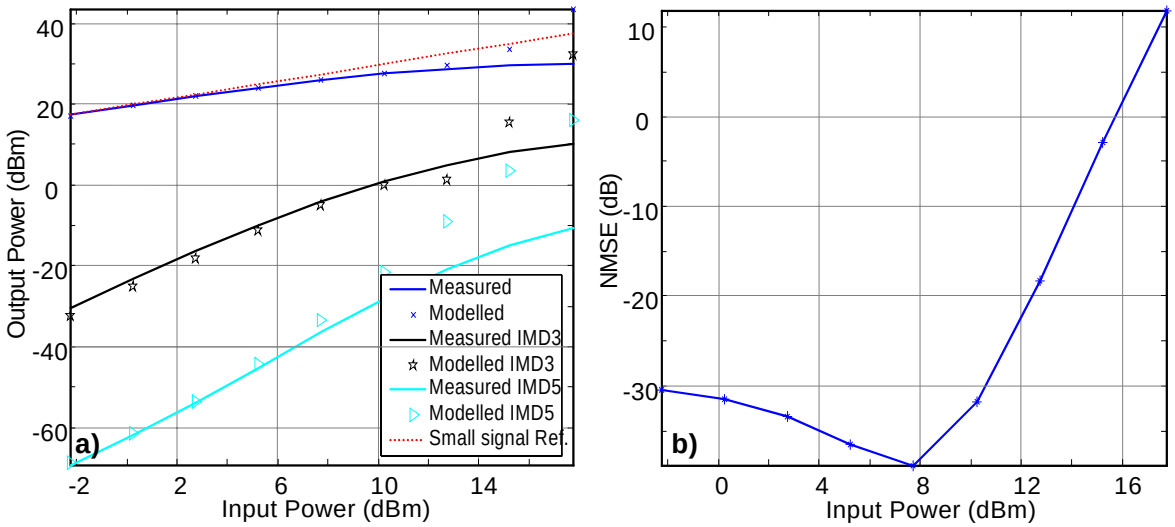


Figure 4.32 – a) Model results on output power, IMD3 and IMD5 with input signal power sweep. b) Normalized Mean Square Error of the model for the Wiener-Hammerstein System.

4.5 Modelling a Nonlinear Amplifier With Memory Effects Caused by the Bias Circuitry

In this section the modelling approach is validated in a more realistic power amplifier. Figure 4.33 shows the block diagram of the amplifier being considered. The feedback path in this configuration is responsible for a new type of memory that none of the previous systems can mimic. This memory does not appear if the system is operated in the small signal condition, this is why this memory effect is usually called nonlinear memory.

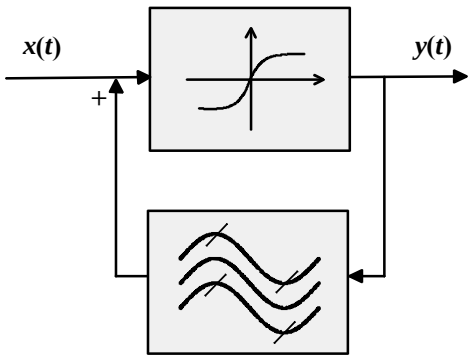


Figure 4.33 – Schematic representation of the Hammerstein model configuration.

Figure 4.34 presents the circuit schematic of the block diagram considered. If the L2 inductor value is properly chosen than it creates considerable reactive base band impedance that varies within the input signal bandwidth [4.4]. This schematic was implemented in ADS, and then properly tuned to present nonlinear memory effects.

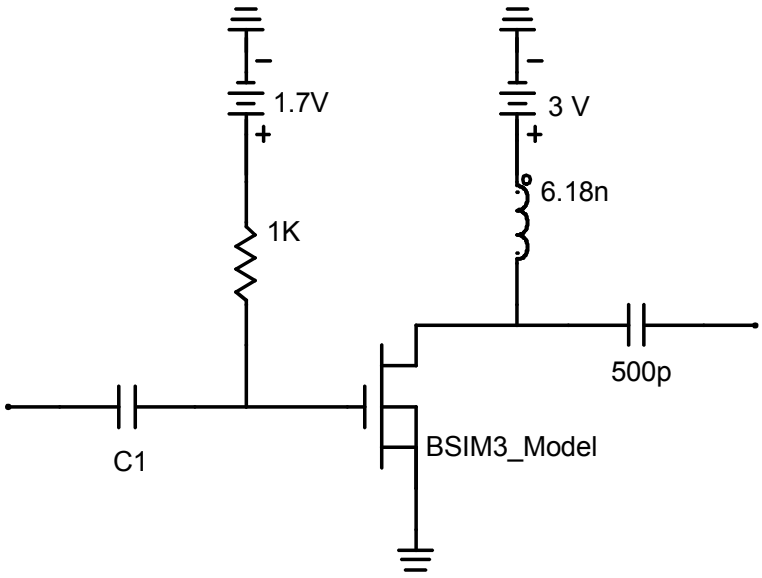


Figure 4.34 – Circuit diagram of the amplifier being modelled.

As in the previous sections, the main characteristics of the system under test are plotted in the next figures. The small-signal S_{21} parameters are represented bellow:

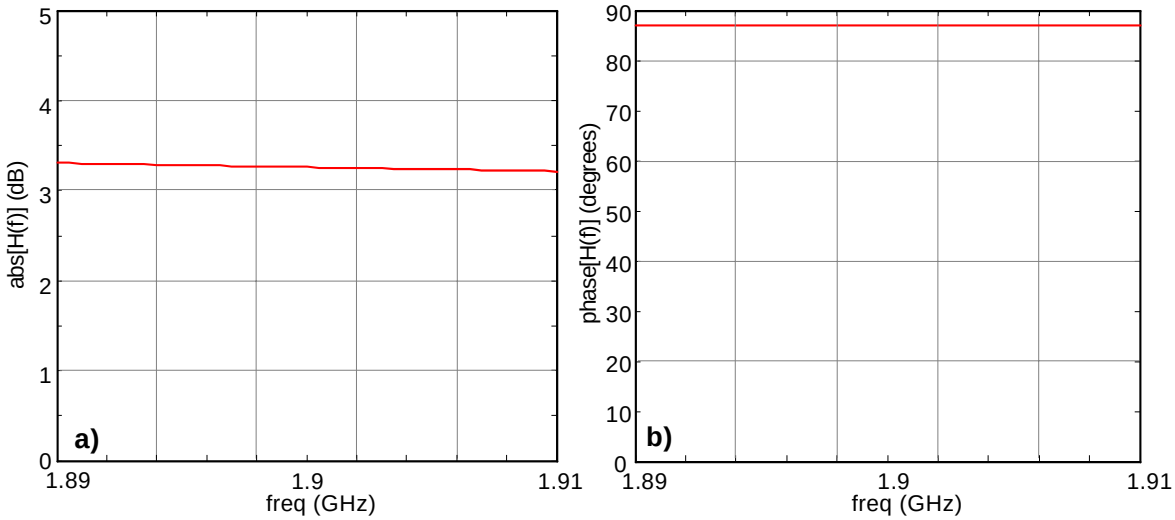


Figure 4.35 –Small signal gain variation with frequency of the simulated PA considered.

This figure presents a very small variation on the S_{21} from which we might assume the small signal flat frequency response of the system. Figure 4.36a) presents the one tone

input/output power curve of the device showing this way its static nonlinear properties. The instantaneous input/output power measured with a two tone signal is shown in Figure 4.36b) and demonstrates its nonlinear dynamics. These are mainly visible for high input powers.

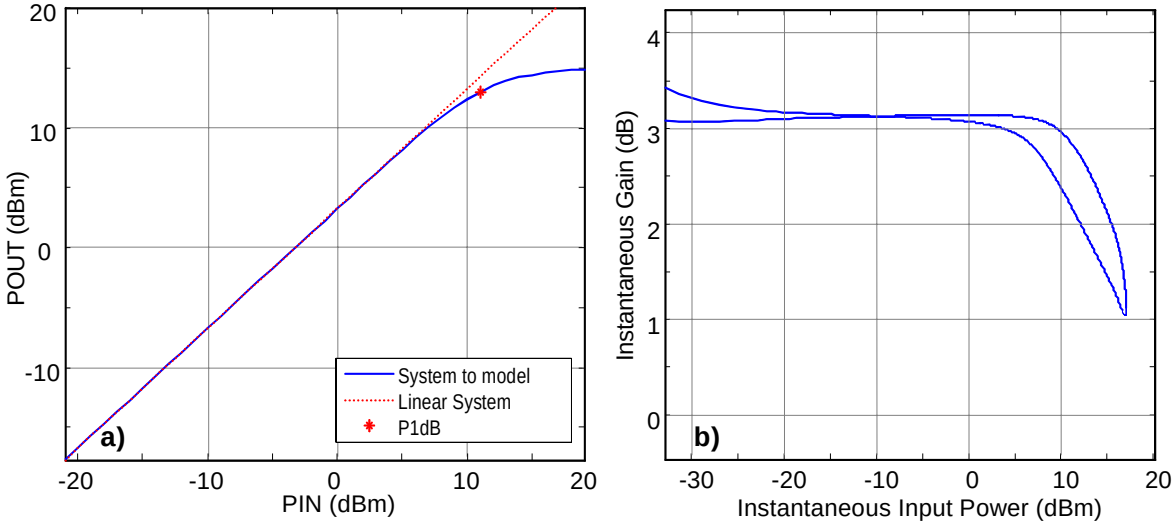


Figure 4.36 – a) One tone average P_{IN}/P_{OUT} of the system of Figure 4.15. b) Instantaneous two-tone gain curve for the simulated PA configuration.

After the main system characterisation, some modelling results are now presented. As in the previous examples the first model results are the static and dynamic comparison between measured and modelled output signals presented in Figure 4.37.

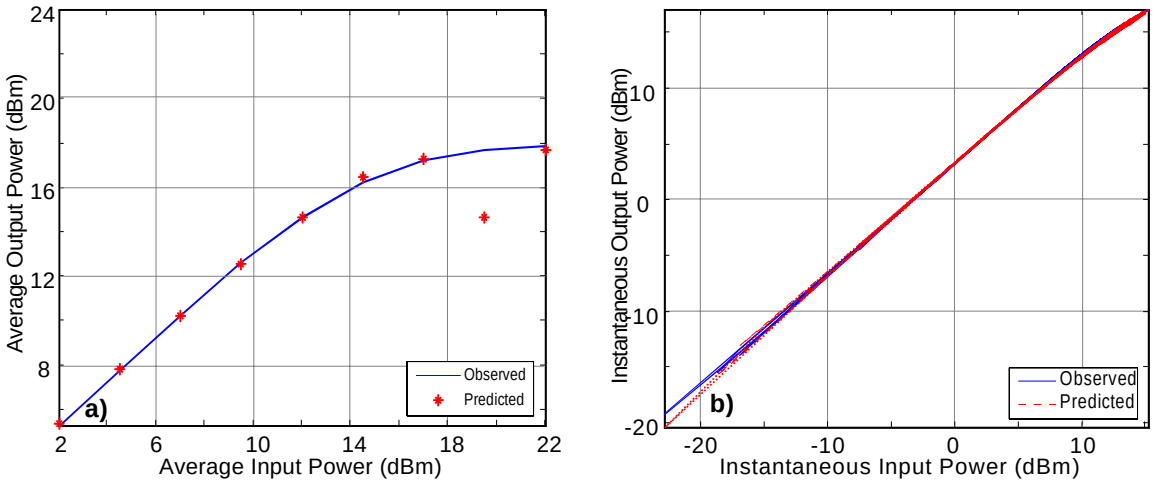


Figure 4.37 – a) Observed and Predicted input/output one tone average power. b) Comparison between Modelled and Measured WCDMA instantaneous output for the simulated PA.

Figure 4.37a) shows, once more, the good approximation results for the small signal regime and the model divergence for high input powers. Figure 4.37b) presents the instantaneous input/output curves; it is visible in the higher input powers a slight

enlargement of the input/output power curve. This enlargement is due to the nonlinear memory effects.

Figure 4.38a) shows the complex time domain magnitude comparison between measured and modelled WCDMA envelope signals. The corresponding frequency spectra comparison is in Figure 4.38b). Good approximation results were obtained.

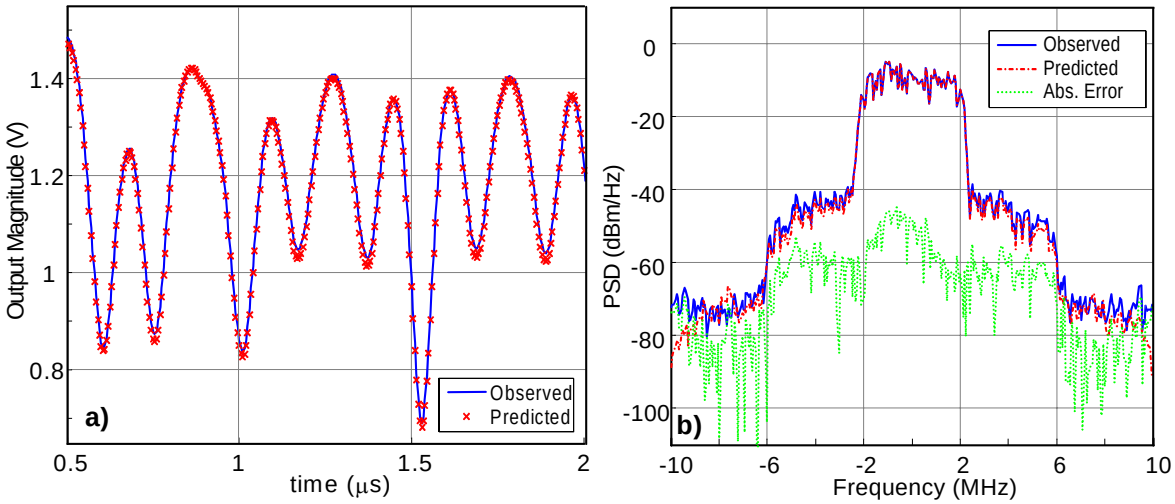


Figure 4.38 – a) Complex envelope time domain comparison between Measured and Modelled WCDMA signals. b) The predicted and observed spectra of WCDMA validation signal.

Figure 4.39a) shows the comparison between measured and modelled power of the in-band, adjacent and alternate channels. In this situation, the measured adjacent and alternate channel powers are quite noisy.

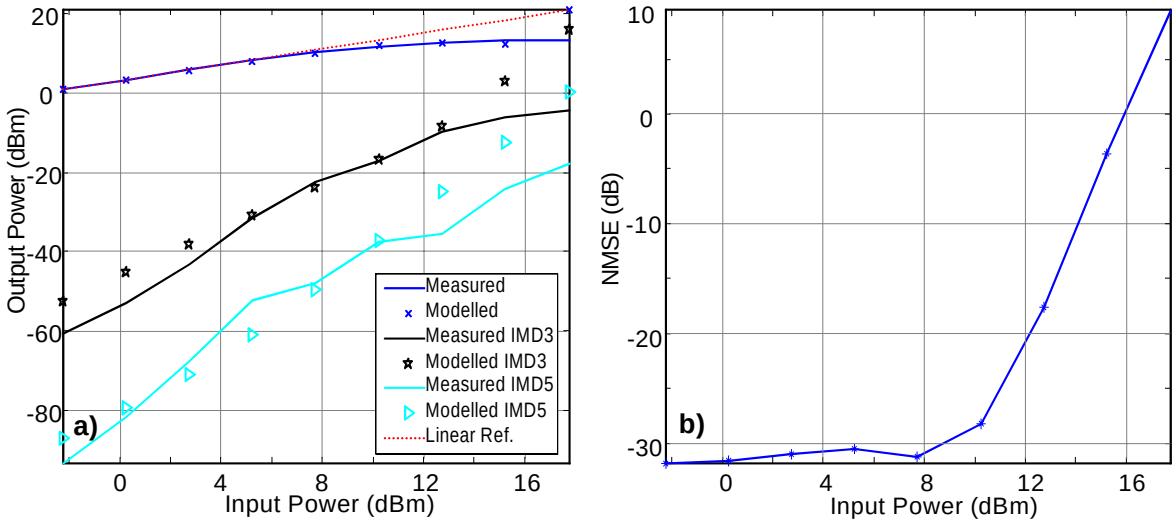


Figure 4.39 – a) Model results on output power, IMD3 and IMD5 with input signal power sweep. b) Normalized Mean Square Error of the model response for several input powers.

4.6 Modelling a Real Power Amplifier

In order to demonstrate the application of this modelling approach on a real power amplifier two different tests have been made. Profiting from the collaboration with other European universities and companies within the Network of Excellence TARGET – Top Amplifier Research Group in a European Team, we had access to two comparable amplifiers designed to study the impact of DC-bias network effects on the amplifier nonlinear memory characteristics. This way the model performance could be tested on one amplifier without significant memory effects and also in an amplifier specifically designed to present nonlinear memory effects [4.5,4.6].

The next two subsections present the main characterization of each device as well as the modelling performance obtained.

4.6.1 Nonlinear Memoryless Amplifier

Figure 4.40 presents the circuit topology of the considered memoryless amplifier. The circuit component values were chosen to obtain a flat frequency response near the operation band. This amplifier was modelled for a CDMA2000 signal application centered at 950 MHz.

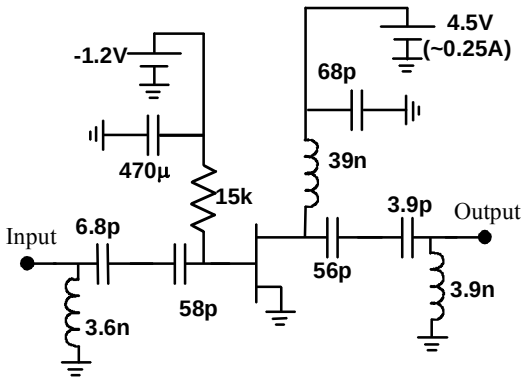


Figure 4.40 – Schematic representation of the circuit of the memoryless amplifier.

For this system the small-signal S_{21} parameters are represented in Figure 4.41. The analysis of this figure indicates that it is a good assumption to consider a flat response in a band of nearly 5 MHz around 950 MHz where the small signal S_{21} varies less than 0.1 dB.

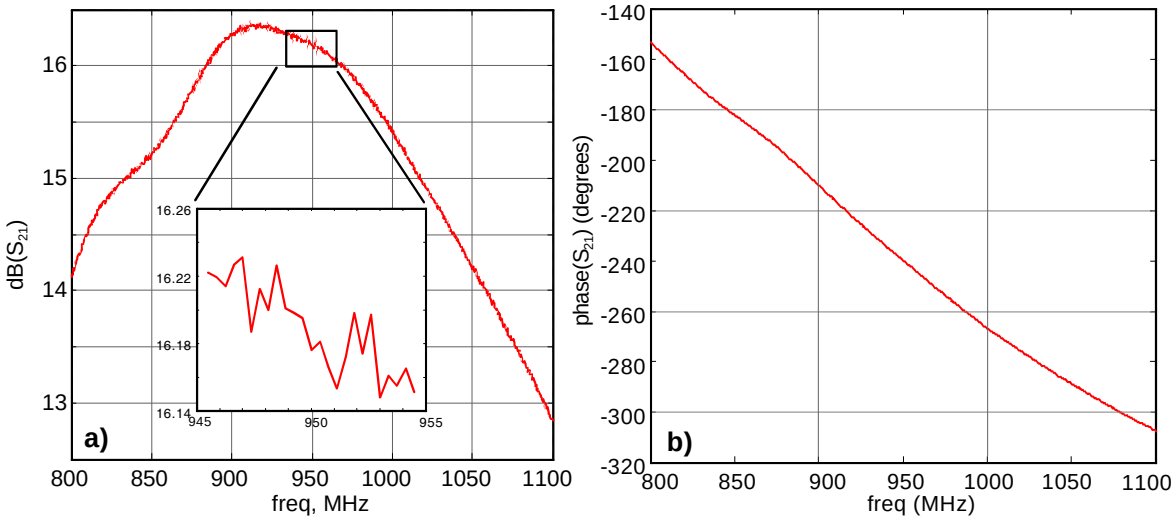


Figure 4.41 –Small signal gain variation with frequency of the Wiener system considered.

To characterize the nonlinear behaviour of this amplifier, a one tone input/output power sweep was made and the results are shown in the Figure 4.42. A fifth order model, with seven time delays was extracted for an average input power of 6 dBm which is approximately the 1 dB compression point.

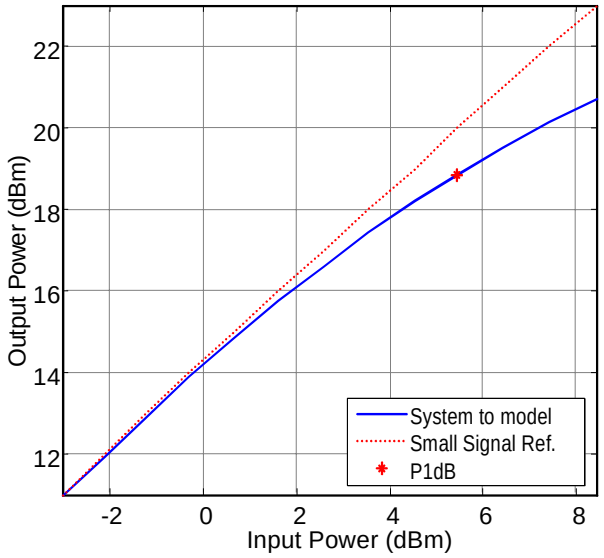


Figure 4.42 – One tone average P_{IN}/P_{OUT} of the memoryless amplifier.

The cross validation results are shown in the next two figures. Figure 4.43 shows a comparison between the envelope magnitudes of the amplifier response measured and modelled. It is seen some large deviation especially for abrupt transitions and high peaks.

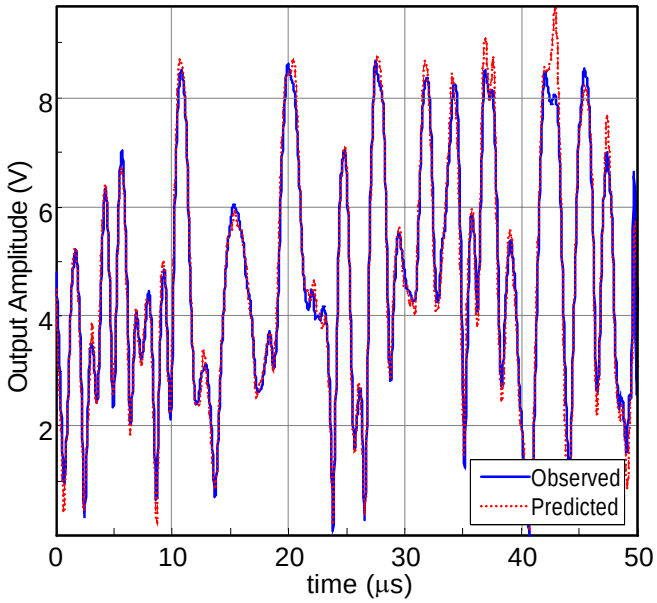


Figure 4.43 - Time domain magnitude comparison between Measured and Modelled multitone signals.

At this operation power level the PA is already clipped as Figure 4.43 demonstrates, and so a polynomial model has difficulties describing the system's output response (Compare with Figure 4.45).

The analysis of the spectra comparison in Figure 4.44 shows a larger difference between measured and modelled output signals for lower output PSDs. Other issue that might influence this modelling problem are the different statistical properties (PSD, PAR, etc) of the multisine used for model extraction and of the CDMA2000 sequence used for validation.

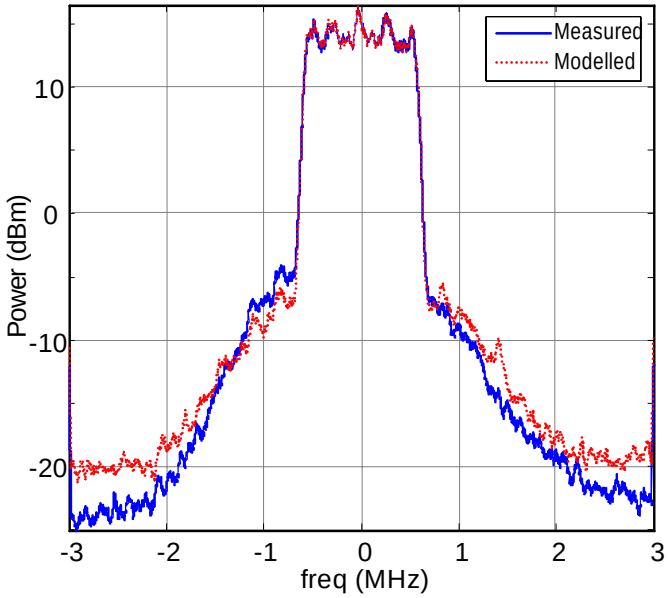


Figure 4.44 – The predicted and observed spectra of CDMA validation signal.

The instantaneous input/output power measured and modelled comparison is shown in Figure 4.45. In this figure a reasonable similarity is observed between both curves. The polynomial model divergence starts to be visible in the high power zone.

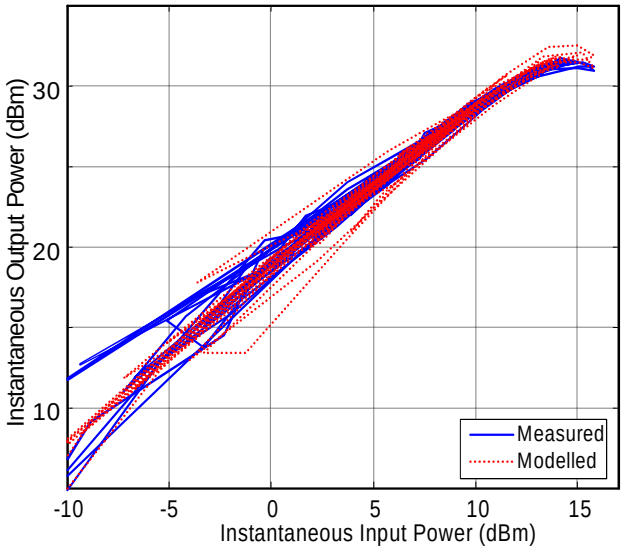


Figure 4.45 –WCDMA signal instantaneous input/output power curve for the memoryless amplifier – Measured and Modelled.

Table 4.1 shows the NMSE obtained for this amplifier for different CDMA2000 signal average input powers. Being slightly worse than the results obtained with the model extracted for simulated results, the NMSE values above 20 dB were still a reasonable achievement for a behavioural model obtained from real measurements.

Table 4.1 – Normalized mean square error for different CDMA2000 input signal powers.

Input Power (dBm)	NMSE (dB)
4	-22.3
6	-25
9	-23.45

4.6.2 Nonlinear Dynamic Amplifier

A different topology of the bias networks, shown in Figure 4.46, introduces nonlinear memory effects. The bias resonance, specifically added to create a sharp frequency

response within the input signal bandwidth, is the main cause of this nonlinear dynamic behaviour.

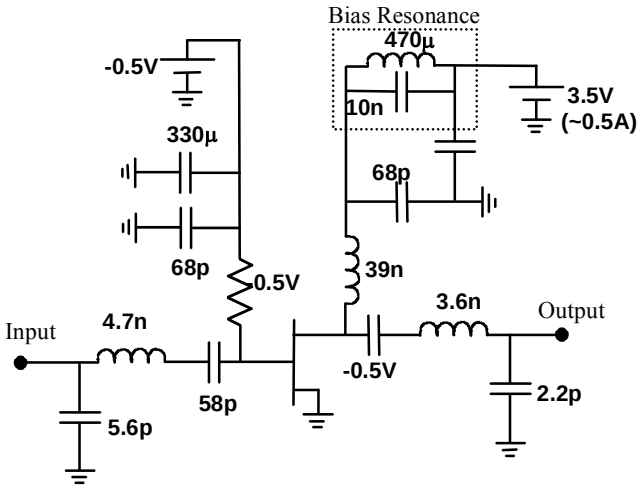


Figure 4.46 – Schematic representation of the circuit of the nonlinear dynamic amplifier.

The model for this amplifier was extracted in the same conditions of the previous one. Once more, the small-signal S_{21} parameters of the system being studied are presented.

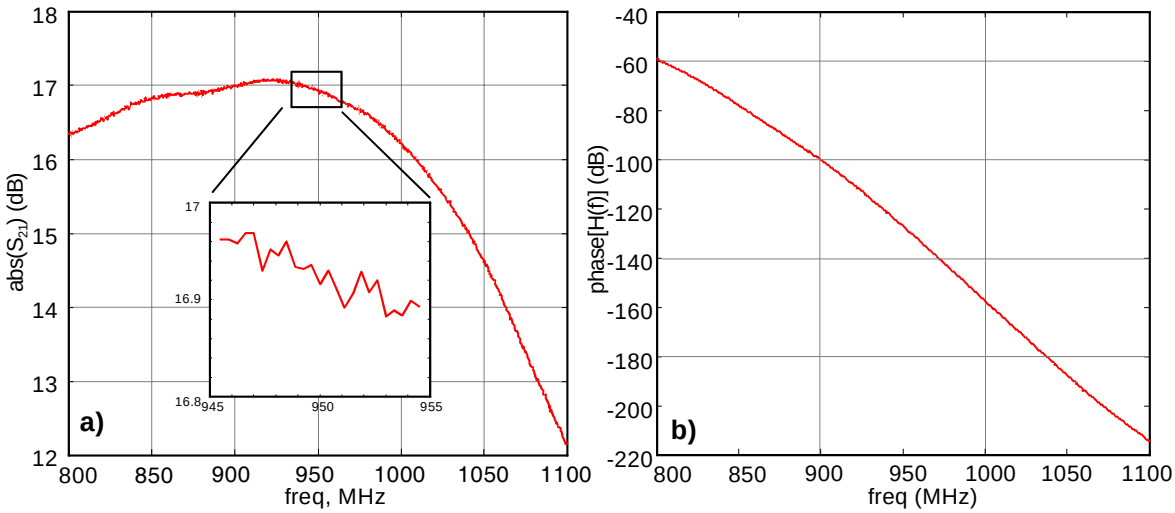


Figure 4.47 –Small signal gain variation with frequency of the nonlinear dynamic amplifier.

The nonlinear static characterization is shown in Figure 4.48. This input/output power curve was measured using a single tone signal.

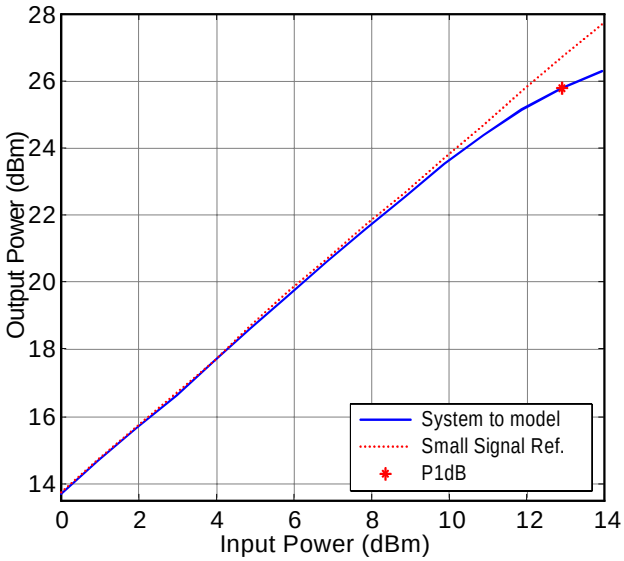


Figure 4.48 – One tone average P_{IN}/P_{OUT} of the dynamic amplifier.

The next two figures show the cross validation results obtained. In Figure 4.49 the envelope magnitude of the CDMA2000 signal used is plotted. A good comparison is seen between the modelled and predicted signals. The frequency domain equivalent results are shown in Figure 4.50.

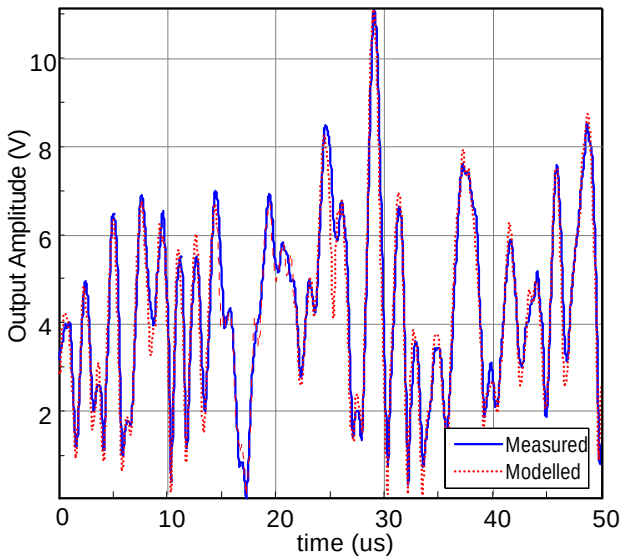


Figure 4.49 – Time domain Magnitude comparison between Measured and Modelled CDMA signals.

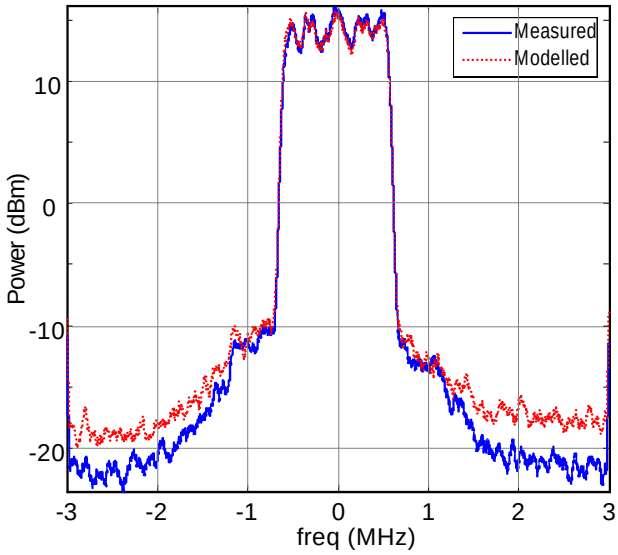


Figure 4.50 – The predicted and observed spectra of CDMA validation signal.

The instantaneous input/output plot obtained with the CDMA2000 signal is shown in Figure 4.51. This figure denotes some difficulties of the model in accurately modelling the nonlinear memory component. This difficulty might be a consequence of the measurement error, because the model worked just fine in the simulated amplifier with nonlinear memory (Figure 4.37b). Another possible explanation for the model's bad performance modelling the nonlinear memory on this situation might be an insufficient discretization of the spectra to accurately capture the nonlinear frequency response imposed by the bias resonance shown in Figure 4.46.

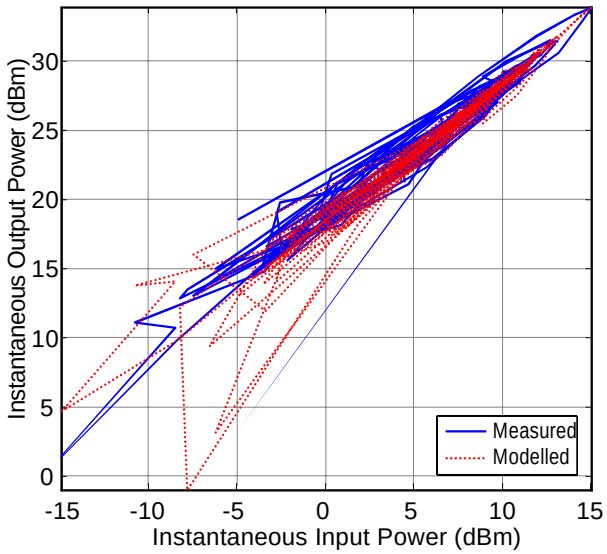


Figure 4.51 – Instantaneous CDMA2000 input/output power curve for the nonlinear dynamic amplifier – Measured and Modelled.

Table 4.2 presents the NMSE evolution with the input power sweep. The poor results obtained in this situation are probably due to the insufficient spectral discretization to represent the nonlinear system dynamics'. The choice of a small number of tones in this situation was made to allow the extraction of a fifth order model – since the number of coefficients grows with L^N (memory length to the power of nonlinear order) to increase the order and keep a reasonable number of coefficients the number of time delays has to be small. However these poor results appear to indicate that, in this case, probably a smaller order model with more time delays would perform much better, but this experiment was not possible to perform since the amplifiers were measured on a round-robin basis and were not available to repeat the measurements when this issue was detected.

This table points out, again, the local characteristics of the model.

Table 4.2 – Normalized mean square error for different CDMA2000 input signal powers.

Input Power (dBm)	NMSE (dB)
4	-14.7
6	-17.0
9	-19.4
11	-14.7

4.7 Conclusions

In this chapter an extensive set of validation tests were performed to demonstrate the model capabilities. The model operation and performance was demonstrated both on the regions where it performs well and on the regions where the divergence is obvious.

It was shown that this model has the main behavioural characteristics of a "conventional" Volterra series: it behaves well for small signal (bellow the extraction point), and diverges (rapidly) for the large signal operation. One improvement of this model is that since it is a local approximation – in opposition to the small signal approach of the conventional Volterra series – its extraction point can be shifted to extract a model more devoted for a given application scenario.

The modelling capabilities were demonstrated even in difficult situations as a nonlinear amplifier with nonlinear memory effects. Good results were obtained in the case of the simulated amplifier.

For the real dynamic amplifier, presented in section 4.6.2, the number of time delays considered and/or the sample frequency was unable to capture the system's dynamic behaviour. However, the modelling of this amplifier was limited by the number of coefficients that can be treated (which limited the number of time delays) and also by the bandwidth of the measurement equipment used which did not allow a high sampling rate.

4.8 References

[4.1] J. C. Pedro, N. B. Carvalho and P.M. Lavrador, "Modeling nonlinear behavior of bandpass memoryless and dynamic systems," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, U.S.A, 2003, pp. 2133-2136.

[4.2] Agilent, "Advanced Design System," 2005 ed.

[4.3] P. M. Cabral, J. C. Pedro and N. B. Carvalho, "Dynamic AM-AM and AM-PM Behaviour in Microwave PA Circuits", *Asia and Pacific Microwave Conference Proc.,* Suzhou, China, Dec. 2005, pp. 2386-2389.

[4.4] N. B. Carvalho, J. C. Pedro, "A Comprehensive Explanation of Distortion Sideband Asymmetries," *IEEE Trans. on Microwave Theory and Techniques*, vol. 50, Issue: 9, pp. 2090 – 2101, Sept. 2002

[4.5] TARGET Internal Report 2.2.E.2.5 on the Joint Investigation Projects of WP 2.2.E.2 Amplifier Modelling; edited by Anthony Goacher and Dominique Schreurs, disseminated between WP members.

[4.6] C. Fager, P. M. Lavrador, M. Myśliński, A. Zhu, H. M. Nemati1, J.C. Pedro and D. Schreurs, "System Level Modelling of RF Power Amplifiers Based on Large-Signal Measurements", submitted to EuMW 2007.

5. Noise and Distortion Figure: An Outcome of Cross-Correlation Identification.

5.1 Introduction

In the previous Chapters an orthogonal model was formulated and validated making use of output signal separate components identification. The separable identification was based on the definition of cross-correlation between the output and a set of particularly selected input combinations.

If the input signal itself is used, and the output part correlated with the input is computed, than it is possible to separate the output signal and non-signal components. After the identification of these two components, an easy metric that accounts for the signal degradation imposed by the system can be achieved.

In this Chapter, by using white Gaussian noise as the standard excitation, and cross-correlation techniques, it is shown how the Best Linear Approximator (BLA), can be determined for general memoryless and dynamic nonlinear Volterra systems, this way

allowing the identification of the desired signal and noise components involved in the system's output. Considering these results, a new Figure of Merit (FoM), named Noise and Distortion Figure (NDF), is proposed to evaluate the total signal degradation due to noise and distortion, simultaneously. It starts with a motivation to the usefulness of the new FoM based on the currently existing ones. After this discussion, the usual Noise Figure (NF) is revisited, and a formal discussion on the output separation of signal and noise for a system modeled by a Volterra Series is presented. In order to validate the derived closed form expressions for the newly defined NDF, a time domain simulation was performed for a typical dynamic nonlinear system, of third order, and the results compared with the proposed theoretical values.

5.2 Current Figures of Merit for Signal Degradation Evaluation

Signal to noise ratio, SNR, in real communication systems can be severely degraded when signals are processed by nonlinear components. That degradation is normally attributed to two different impairments: linear additive noise and nonlinear distortion [5.1].

In order to account for the additive noise, the figure of merit Noise Figure, NF, is normally used, while the 3rd order Intercept Point, IP_3, can be made to play a correspondent role for nonlinear distortion degradation.

Unfortunately, until now the complex nature of nonlinear distortion has prevented the integration of these two SNR degradation figures, forcing the design engineer to evaluate any link budget in two different steps: looking for the small amplitude signal limitations determined by additive noise, and its high level end imposed by nonlinear distortion. Only by taking into account those two perturbation causes, he can maximize the communication systems' dynamic range.

In an effort to understand the relation between these two signal perturbation figures of merit, A. Geens and Y. Rolain have detected in [5.2] some problems when measuring NF in the presence of nonlinearities and proposed a new noise figure to circumvent those problems. Nevertheless, and due to the excitation signals that were used, the results obtained with this new formulation of the NF can be disastrous, as it predicts certain zones of improvement in the output signal to noise ratio (SNR_o), an obviously impossible outcome in practical situations. Furthermore, this work restricted its analysis to

memoryless nonlinear systems, which constitutes a severe limitation if applied to modern wide band wireless components that are known to exhibit non-negligible nonlinear memory effects [5.3,5.4].

One of the first and most important difficulties imposed by nonlinear distortion analysis is its dependence on the type of signal excitation. That issue, for long time recognized in the nonlinear systems' identification field [5.5,5.6], demands for a careful selection of a convenient signal class.

Although in the past, RF and microwave engineers had a propensity to represent their telecommunication signals by a pure sinusoid, it is already known that such class of signals is totally inadequate. In fact, it can not represent any real signal of non null bandwidth, amplitude varying envelope and the random behavior associated to information. Although the two-tone has been also widely adopted for nonlinear distortion testing, it still suffers from the fact that it only involves a sinusoidal envelope of deterministic behavior.. A much better signal class used to represent real communication signals that does not suffer from any of these drawbacks is the band-limited White Gaussian Noise, which will be therefore adopted for the present definition of a new figure of merit intended to be a metric of SNR degradation in presence of additive noise and nonlinear distortion.

The second issue worth of discussion is the separation of the system's output into its signal and distortion components. A useful criterion should be to use the same separation undertaken in modern wireless receivers as it would immediately lead to practically significant transmission quality figures as bit-error-rate. So, in that sense, it is considered as signal everything that contains information possibly processed by a linear dynamic operator, and as distortion any remaining part. This way it is possible to classify as signal the outcome of the so-called Best Linear Approximation, BLA, [5.6] which governs the linear behavior of the output signal versus the input excitation, and then use cross-correlation to uniquely identify it. The dependence of the system's Best Linear Approximation (gain in a memoryless system) on the input has been already discussed in the 60's [5.7]. With those assumptions in mind, it was then possible to correctly divide the output useful signal from the noise distortion, and then quantify the signal to noise and distortion ratio, $SINAD_0$, at the output.

5.3 Nonlinear Noise Figure Revisited

An important figure of merit used to measure the degradation of signal quality between input and output is the NF which relates the Signal to Noise Ratio at the input (SNRi) to the Signal to Noise Ratio at the output. A. Geens and Y. Rolain [5.2] have proved that the presence of nonlinear distortion influences the measured NF value, and proposed a new setup for measuring NF using a single tone as a test signal. In this work an input composed by the sum of a single tone, of amplitude A, and band-limited, white, zero-mean Gaussian noise with single sided PSD equal to N_0 was used. To demonstrate the impact of the nonlinearity on the NF, a noiseless system is considered, and so a NF = 1 should be obtained if the system were also linear. Using this approach, they reached the following expression for the NF:

$$NF = \frac{SNR_{in}}{SNR_{out}} = \frac{G + \frac{27}{8}\alpha^2 A^4 - 3\alpha\sqrt{G}A^2}{G + \frac{9}{16}\alpha^2 A^4 - \frac{3}{2}\alpha\sqrt{G}A^2} \tag{5.1}$$

where G is the linear power gain, α is a third order voltage gain and A the input tone amplitude. A closer look into expression (5.1) reveals that there are certain zones of input signal voltage in which the NF can be smaller than one, as shown in Figure 5.1.

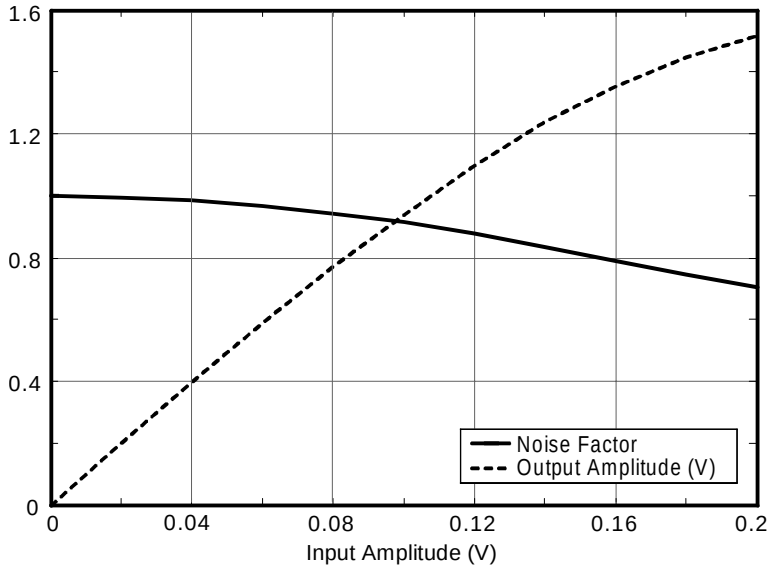


Figure 5.1 – The NF proposed in [5.2] variation with the input power, for a system with parameters G = 100, α = 60, as indicated in the paper.

This result is strange since it indicates that the system can, in fact, improve SNR from the input to the output, in a certain sense eliminating input noise.

A closer look into this theoretical result shows that the gain in the SNR, verified in these conditions, is caused by the different compression imposed to each signal: a sinusoid and white Gaussian noise. Actually, it is known [5.8] that when two different signals excite a nonlinear system, in which one is of much larger amplitude than the other, the compression of the smaller one is mainly determined by the level of the strong signal. In this case, the sinusoidal signal is the dominant component therefore determining a greater compression to the noise. In fact, the relation between the output sinusoid and noise will be improved due to the extra compression imposed to noise level. The referred NF characteristics associated to this approach can be traced to the use of a single sinusoid as the input signal. Actually, there is no input noise perturbing the signal, since the signal has a null spectral bandwidth, and thus there is no noise power inside the signal bandwidth. Additionally, since this test signal has a constant envelope, it is also unable to generate uncorrelated nonlinear distortion, also known as nonlinear distortion noise [5.6].

A more appropriate alternative would be the use of a test signal similar to a real communications signal, for example Gaussian noise, since it has nonzero bandwidth allowing the inclusion of effective additive noise and uncorrelated nonlinear distortion effects. Beyond that it has statistical properties similar to the ones of real signals.

5.4 Signal and Noise Identification

Despite the advantages of using Gaussian signals pointed out in the last section, there are also several difficulties associated with the separation between signal and noise components. In this case, the signal and noise share the same spectral positions obviating any straightforward separation in the frequency domain. Moreover, the signal component may be several orders of magnitude higher than the noise level.

A physical meaning solution, often used because of its practical interest, is to consider as signal the output component correlated with the input, as is usual in conventional rake receivers. This result is supported by Bussgang's Theorem [5.9]. In Figure 5.2 it is shown a geometric illustration of this operation. The projection of the

output (vector v) - which has correlated (collinear) components with the input and other ones uncorrelated (orthogonal) - onto the input signal component (vector u) is calculated using the input/output correlation and the power of input and output signals. That projection is the output signal component (vector w).

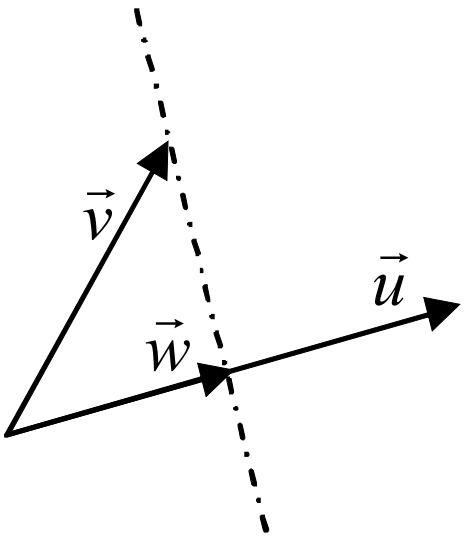


Figure 5.2 – Geometric representation of the method used to determine the output signal component.

One way to obtain that correlated component is to use the BLA, which is defined as the linear transfer function that is the best approximation to the nonlinear system in a least squares sense [5.6]. In the frequency domain, it can be given by:

$$H_L(\omega) = \frac{S_{xy}(\omega)}{S_{xx}(\omega)} \tag{5.2}$$

where $S_{xy}(\omega)$ is the cross spectral density of the input and output signals and $S_{xx}(\omega)$ is the input spectral density function that can be calculated as the Fourier Transform of the cross-correlation and autocorrelation functions respectively:

$$R_{xy}(\sigma) = \overline{x(t)y(t+\sigma)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)y(t+\sigma)dt \tag{5.3a}$$

$$R_{xx}(\sigma) = \overline{x(t)x(t+\sigma)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t+\sigma)dt \tag{5.3b}$$

Having defined this way the signal component, it can thus be considered as nonlinear noise all the remaining output components. Part of this distortion noise is irrelevant as it falls out-of-band (i.e., around DC and the carrier harmonics) and thus can be eliminated by proper filtering. The remaining in-band noise is present in the co- and adjacent-channels. To compute the SINAD one must consider as relevant noise only the co-channel part.

The first approach considered is to calculate the output signal component in the case where a memoryless nonlinearity is modeled by a power series.

5.4.1 BLA Calculation for a Memoryless Nonlinearity.

A Gaussian signal, $x(t)$, will be considered as input, so that the output is given by:

$$y(t) = \sum_{n=1}^{N} \alpha_n x(t)^n \tag{5.4}$$

In order to obtain the output signal component, the input/output cross-correlation must be evaluated. Applying the definition of expression (5.3) and the properties of averaging Gaussian random variables [5.5] $R_{xy}(\tau)$ will be

$$\begin{aligned} R_{xy}(\sigma) = {} & \alpha_1 R_{xx}(\sigma) + 3\alpha_3 R_{xx}(0)R_{xx}(\sigma) \\ & + 15\alpha_5 R_{xx}(0)^2 R_{xx}(\sigma) + \ldots \end{aligned} \tag{5.5}$$

which can be written in a generalized form as

$$R_{xy}(\sigma) = R_{xx}(\sigma) \cdot \sum_{n\ odd}^{N} \alpha_n \frac{(n+1)!}{\left(\frac{n+1}{2}\right)! 2^{\left(\frac{n+1}{2}\right)}} R_{xx}(0)^{\left(\frac{n-1}{2}\right)} \tag{5.6}$$

Expression (5.6) presents a general result for input/output cross correlation of a memoryless nonlinear system modeled by a *N*'th order polynomial. This expression indicates that correlation only exists between output odd order terms and the input, since it is known that the average of the product of a number of Gaussian random variables is only non zero if that number is even.

With equation (5.2) and equation (5.6) the linear transfer function (or gain) of a memoryless nonlinearity modeled by a polynomial can be directly expressed as:

$$H_L(\omega) = \sum_{n\ odd}^{N} \alpha_n \frac{(n+1)!}{\left(\frac{n+1}{2}\right)!2^{\left(\frac{n+1}{2}\right)}} R_{xx}(0)^{\left(\frac{n-1}{2}\right)} \tag{5.7}$$

This expression states that the Best Linear Approximation is not only dependent on the system parameters, α_i, but also on the input signal characteristics, namely its power.

5.4.2 BLA Calculation for a Nonlinear System with Memory

A nonlinear system that presents memory, but is sufficiently well behaved so that it can be described by a Volterra Series is now addressed. Although conceptually similar to the memoryless case, this problem is significantly more difficult to treat analytically. The derivation process starts by writing the analytical expression for the output (5.8)

$$\begin{aligned} y(t) = & \int_{-\infty}^{+\infty} h_1(\tau)x(t-\tau)d\tau + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1 d\tau_2 + \\ & \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h_3(\tau_1,\tau_2,\tau_3)x(t-\tau_1)x(t-\tau_2)x(t-\tau_3)d\tau_1 d\tau_2 d\tau_3 \\ & + \quad \cdots \end{aligned} \tag{5.8}$$

or, in a general form:

$$y(t) = \sum_{n=1}^{N} \int_{-\infty}^{+\infty} \cdots \int_{n\ fold} \int_{-\infty}^{+\infty} h_n(\tau_1, \tau_2, \ldots, \tau_n) x(t-\tau_1) x(t-\tau_2) \cdots x(t-\tau_n) d\tau_1 d\tau_2 \cdots d\tau_n \tag{5.9}$$

Once again, using the definition (5.3), the cross-correlation between input and output can be calculated. Considering a Gaussian random signal $x(t)$, and y(t) given by (5.9), $R_{xy}(\tau)$ in expression (5.10) is obtained.

$$\begin{aligned} R_{xy}(\sigma) &= \sum_{n} R_{xy_n}(\sigma) \\ &= \int_{-\infty}^{+\infty} h_1(\tau) R_{xx}(\sigma-\tau) d\tau + 3 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_3(\tau_1, \tau_2, \tau_3) R_{xx}(\sigma-\tau_1) R_{xx}(\tau_2-\tau_3) d\tau_1 d\tau_2 d\tau_3 + \cdots \\ &= \sum_{n\ odd}^{N} \frac{(n+1)!}{\left(\frac{n+1}{2}\right)! 2^{\left(\frac{n+1}{2}\right)}} \int_{-\infty}^{+\infty} \cdots \int_{n\ fold} \int_{-\infty}^{+\infty} h_n(\tau_1, \ldots, \tau_n) R_{xx}(\sigma-\tau_1) \cdot \ldots \cdot R_{xx}(\tau_{n-1}-\tau_n) d\tau_1 \ldots d\tau_n \end{aligned} \tag{5.10}$$

The same procedure of the last section will now be used to determine the Best Linear Approximation. The Fourier Transform of equation (5.10) is computed and then (5.2) is used to find $H_L(\omega)$. In order to find the Fourier Transform of (5.10) $h_n(\tau_1,\ldots,\tau_n)$ is written as a function of the n-dimensional inverse Fourier Transform of $H_n(\omega_1,\ldots, \omega_n)$. Then, changing the order of integration between ω's and τ's and using some properties of the Fourier Transform, the following results is reached (5.11):

$$H_L(\omega) = \sum_{n\ odd}^{N} \frac{(n+1)!}{\left(\frac{n+1}{2}\right)! 2^{\left(\frac{n+1}{2}\right)}} \int_{-\infty}^{+\infty} \cdots \int_{\frac{n-1}{2}\ fold} \int_{-\infty}^{+\infty} H_n\left(\omega, \omega_1, -\omega_1, \ldots, \omega_{\frac{n-1}{2}}, -\omega_{\frac{n-1}{2}}\right) S_{xx}(\omega_1) \ldots S_{xx}\left(\omega_{\frac{n-1}{2}}\right) d\omega_1 \ldots d\omega_{\frac{n-1}{2}} \tag{5.11}$$

This expression (which to the best of the authors' knowledge is new) gives the Best Linear Approximation of a nonlinear dynamic system modelled by a Volterra Series of order N when subject to a Gaussian input signal. It states that the Best Linear Approximation is dependent, not only on the system parameters and the even order moments of the input (the integrated power), but also on the stimulus' power spectral

density [the shape of $S_{xx}(\omega)$]. The main interest of expression (5.11) resides on the fact that the BLA varies with the input signal power spectral density in a way that can be interpreted as if $S_{xx}(\omega)$ were 'weighting' the nth order Volterra kernel. Therefore, $H_L(\omega)$ will be different whenever $S_{xx}(\omega)$ gives more importance to the different parts of the multidimensional frequency response of each of the $H_n(\omega_1,\ldots, \omega_n)$.

5.5 Noise and Distortion Figure

Having developed the theoretical tools to isolate the signal components from the noise components, it is now possible to define a new figure of merit that simultaneously deals with noise and distortion [5.10].

It is well known the relation between NF and Signal to Noise Ratio, i.e. the ratio between signal and noise powers. As a matter of fact, although the NF is frequently referred as the ratio between input and output SNRs, the IEEE adopted formal definition of NF is [5.11]:

$$NF(\omega) = \frac{GN_o(\omega) + N_a(\omega)}{GN_o(\omega)} \tag{5.12}$$

in which, N_o is the output available noise power spectral densities at a given source noise temperature, as seen if the system were noise free, and N_a is the system's added noise, respectively. This way defined, the NF varies with frequency and is thus also named as the *spot noise figure*.

In a nonlinear system the approach described above results incomplete [5.2] because the *SNR* degradation caused by the nonlinear intermodulation noise is not taken in account. Another common figure of merit, which is more useful in the context of nonlinear systems, is the SINAD. According to [5.12], it can be defined as the ratio of signal power spectral density, to noise and distortion power spectral densities, and can thus be expressed as:

$$SINAD_{lin}(\omega) = \frac{S(\omega)}{N(\omega) + D(\omega)} \tag{5.13}$$

where $S(\omega)$, $N(\omega)$ and $D(\omega)$, are, respectively, the Signal, additive Noise and nonlinear distortion power spectral densities.

It was already mentioned above that *NF* can represent the ratio of the SNR_i to the SNR_o. If a similar ratio is evaluated using the SINAD, a figure identical to NF will be found except that it will now also include the distortion impact. Accordingly, it should be called Noise and Distortion Figure (NDF) (5.14).

$$NDF(\omega) \equiv \frac{SINAD_i(\omega)}{SINAD_o(\omega)} = \frac{\dfrac{S_i(\omega)}{N_i(\omega)}}{\dfrac{|H_L(\omega)|^2 \cdot S_i(\omega)}{|H_L(\omega)|^2 \cdot N_i(\omega) + N_a(\omega) + IMD(\omega)}} = \frac{|H_L(\omega)|^2 \cdot N_i(\omega) + N_a(\omega) + IMD(\omega)}{|H_L(\omega)|^2 \cdot N_i(\omega)} \tag{5.14}$$

NDF is thus defined as the ratio of the input SINAD to the output SINAD. In (5.14), $H_L(\omega)$ is the BLA, $N_i(\omega)$ is the input available noise power spectral density at a given reference temperature T_0, $N_a(\omega)$ the power spectral density of the additive noise introduced by the device and seen at the output, and $IMD(\omega)$ the power spectral density of intermodulation distortion delivered to the load. For guaranteeing compatibility with the former IEEE NF definition, these SINADs describe spot frequency values, and so they are defined as the ratio between the spot signal power spectral density function, PSD, to the sum of the spot power spectral density functions of the noise and the distortion, assumed uncorrelated.

At the device's input port, these PSDs refer to the source available powers of the signal and the noise, when the source's equivalent noise internal resistance is at the standard noise temperature (290K). The present NDF definition is therefore assuming that the signal available from the source is undistorted, and so that this situation must be guaranteed if the NDF is to be measured. In fact, what must be guaranteed is that the source's IMD can not generate any appreciable IMD inside the DUT and that its value, when seen at the DUT's output, is much smaller than the one due to the DUT itself.

When SINAD calculations are to be made with this NDF, and the device is isolated, it is naturally expected that the input IMD is zero, since the source can be supposed to

produce an undistorted signal. However, if that is not the case, or if the device is embedded in a chain whose precedent blocks already generate some distortion, then this available distortion PSD should be added to the source available PSD of the signal and the additive noise (since all three are assumed uncorrelated).

Note, however, that, in this latter case, a precise calculation of the total output IMD would require knowledge of the phase of those distortion components, since, being correlated with the ones generated by the DUT, they can not be simply added in power at the output. However, the more usual practical situation is that the precise IMD phase relations are unknown, and no other alternative is left then to rely on a mere absolute value addition. Such a power wise addition would therefore correspond to an average power value as discussed in [5.13].

At the output port, the situation is a little bit more complex, as the DUT's IMD depends on the load termination. So, one has to consider the actual load impedance and define the output PSDs of the signal, IMD and noise as referring to the actual powers delivered to that load. That is, while the IEEE NF definition assumes that the DUT is a system that can be described by an operator whose input variables are the available source PSDs, and the output variables are the available output PSDs, now one has to assume that the system is represented by an operator whose input variables are the input available PSDs while the output variables are the PSDs delivered to the load.

In practice, however, these two distinct definitions will lead to similar values in the vast majority of situations. Indeed, because the output mismatch suffered by the signal is the same as the one suffered by the noise, the ratio of their PSDs (the SNR, which is essential in NF calculation) is an invariant to load impedance. So, a NF defined from output available PSDs, or another one defined from PSDs actually delivered to the load, will only differ if the noise introduced by the load (thermal noise) is significant compared to the noise delivered to that load by the DUT. And, since the equivalent noise temperature of the load and the source are probably the same, this can only happen if a rare situation of a DUT with small gain and very low added noise is to be characterized. Therefore, significant discrepancies between our NDF and previous IEEE NF will only be noticed for DUTs of very small gain and very low added powers of additive noise and distortion. If the load added noise PSD were subtracted from the total noise PSD measured at the load (indeed it can be subtracted because the noise associated to the load is uncorrelated to any

other noise generated in the measurement system), then the NDF and the IEEE NF would again be perfectly consistent.

Finally, since the IEEE NF was originally defined for linear systems, it was always a measure of the system's induced SNR degradation independent of the signal input power. That is so because, keeping the gain constant, the output SNR becomes independent of the input signal power or noise power. On the contrary, NDF is especially useful for nonlinear systems, where the gain and the generated IMD are strongly dependent on the input power. Therefore, it should be of no surprise that the NDF must be defined for a certain input power. In the case of dynamic nonlinear systems, the BLA of (5.11) actually shows that it will even be dependent on the available signal's PSD.

To exemplify the use of NDF, a nonlinear system excited by an input, $x(t)$ composed of a signal $s(t)$ and noise $n(t)$ is considered [$x(t) = s(t) + n(t)$]. The $SINAD_o$ can be calculated if the output signal, noise and distortion components are separately identified. As stated above, those components can be separated using the *BLA*. With input $x(t)$, the output $z(t)$, can be decomposed in:

$$z(t) = h_L(t) * x(t) + y(t) \tag{5.15}$$

where $y(t)$ is uncorrelated with $x(t)$, and has two distinct components: the additive noise introduced by the system and the generated nonlinear distortion. Since the origin of these two components is physically distinct, they are uncorrelated with each other and can thus be added in power. The value of $h_L(t)$ can be determined using the BLA. According to this formulation, the output signal component was found: $h_L(t)*x(t)$. That is, the output signal is the output component that can be obtained with the linear transfer function derived from the input/output cross-correlation. With all these statements the $SINAD_o$ can be written as:

$$SINAD_o(\omega) = \frac{|H_L(\omega)|^2 \cdot S_i(\omega)}{|H_L(\omega)|^2 \cdot N_i(\omega) + N_a(\omega) + IMD(\omega)} \tag{5.16}$$

In this expression $S_i(\omega)$, and $N_i(\omega)$ stand for the input signal and input noise power spectral densities, respectively. $H_L(\omega)$ is the frequency domain BLA transfer function,

$N_a(\omega)$ the power spectral density of the DUT's induced additive noise and $IMD(\omega)$ the power spectral density of the stochastic nonlinear intermodulation distortion.

NDF will now be computed for a particular case of the input and for a nonlinear memoryless system, where the signals are flat over a bandwidth B, with power P_s and P_n. as given by (5.17)

$$\begin{aligned} S_{ss}(\omega) &= \begin{cases} P_s / 2B, & -\omega_H \le \omega \le -\omega_L, \omega_L \le \omega \le \omega_H \\ 0 & , elsewhere \end{cases} \\ S_{nn}(\omega) &= \begin{cases} P_n / 2B, & -\omega_H \le \omega \le -\omega_L, \omega_L \le \omega \le \omega_H \\ 0 & , elsewhere \end{cases} \end{aligned} \tag{5.17}$$

The output power spectral density in the fundamental zone may be obtained, replacing (5.17) in (5.4), up to the third order, and can be written as:

$$\begin{aligned} S_{yy}(\omega) = {} & \left[\alpha_1^2 + 6\alpha_1\alpha_3(P_s + P_n) + 9\alpha_3^2(P_s + P_n)^2\right] \cdot (P_s + P_n)/2B \\ & + 6\alpha_3^2\left(-\omega^2 + (\omega_L + \omega_H)\omega + \frac{B^2}{2} - \omega_L\omega_H\right) \\ & \cdot \frac{3}{8B^3}\left(P_s^3 + P_n^3 + 3P_s^2P_n + 3P_s P_n^2\right) \end{aligned} \tag{5.18}$$

Using expression (5.7) the signal components in (5.18) can be identified and isolated from nonlinear distortion. Taking also into account the effect of additive noise, the in-band output SINAD can be obtained as depicted in expression (5.19):

$$SINAD_o(\omega) = \frac{[\alpha_1 + 3\alpha_3(P_s + P_n)]^2 \, P_s/2B}{[\alpha_1 + 3\alpha_3(P_s + P_n)]^2 \, P_n/2B + 6\alpha_3^2\left(-\omega^2 + (\omega_L + \omega_H)\omega + \frac{B^2}{2} - \omega_L\omega_H\right) \cdot 3 \cdot (P_s + P_n)^3/8B^3 + N_a} \tag{5.19}$$

Using, expression (5.14) and (5.19), in-band $NDF(\omega)$ for this case is given by (5.20)

$$NDF(\omega) = \frac{[\alpha_1 + 3\alpha_3(P_s + P_n)]^2 P_n + \frac{9}{2}\alpha_3^2\left(-\omega^2 + (\omega_L + \omega_H)\omega + \frac{B^2}{2} - \omega_L\omega_H\right)(P_s + P_n)^3 / B^2 + N_a}{[\alpha_1 + 3\alpha_3(P_s + P_n)]^2 P_n} \tag{5.20}$$

where it can be seen that the NDF assumes a parabolic shape inside the band. That is due to the triple convolution of the bandpass signal used in this example.

This expression will not tend to expression (5.1) since the statistical properties of Gaussian noise (even when the bandwidth is narrow) are different from a single sinusoid.

In the case of a nonlinear system with memory, the process is much more laborious but follows exactly the same procedure. First the BLA is calculated using expression (5.11), then the output spectral density function is derived, and these two values are used to compute output noise and distortion.

5.6 Validation of the Theoretical Results

In order to validate the above theory, the NDF of the general system of Figure 5.3 was estimated from time domain numerical simulations and these results compared to the ones directly obtained from expression (5.14). Several tests were performed for different input signals and distinct system configurations.

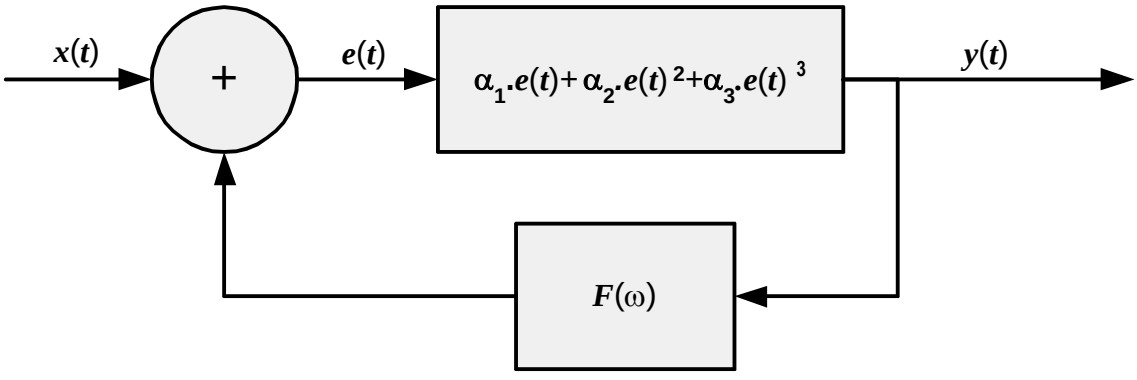


Figure 5.3 – Block Diagram of a general nonlinear bandpass dynamic system.

The Volterra series representation of the system in Figure 3 was obtained in [5.14] and is rewritten here for convenience:

$$H_1(\omega) = \frac{\alpha_1}{D(\omega)} \tag{5.21}$$

and

$$H_3(\omega_1, \omega_2, \omega_3) = \frac{1}{D(\omega_1)\, D(\omega_2)\, D(\omega_3)} \; \frac{1}{D(\omega_1 + \omega_2 + \omega_3)} \left\{ \alpha_3 + \frac{2}{3} \alpha_2^{\,2} \left[\frac{F(\omega_1 + \omega_2)}{D(\omega_1 + \omega_2)} + \frac{F(\omega_1 + \omega_3)}{D(\omega_1 + \omega_3)} + \frac{F(\omega_2 + \omega_3)}{D(\omega_2 + \omega_3)} \right] \right\} \tag{5.22}$$

where $D(\omega)=1-a_1F(\omega)$.

This general system can be set to model both situations presented above, the memoryless nonlinear system and the system with memory.

The memoryless nonlinear system is obtained by eliminating the feedback path, making $F(\omega)=0$, while the system with memory is obtained by proper tuning of the feedback path. In [5.14] it was proved that only an $F(\omega)$ reactive to the base band frequencies can be responsible for the envelope memory effects. Thus, in the dynamic case, $F(\omega)$ was designed to present some reactive behavior at low frequencies.

In order to observe the impact of the input signal spectrum on the BLA, and thus on the NDF, in the dynamic case, this system was simulated recurring to three different input signals.

Since the presented studies are only relative to simulations, the real frequency value is not important. So, the frequency axis of the next figures is normalized, standing for the ratio between the frequency and the sampling frequency.

The simulator block diagram is depicted in Figure 5.4.

5.6.1 NDF Calculation in a Memoryless Situation.

Lets consider first a memoryless nonlinearity ($F(\omega)=0$). Although the theoretical conclusions predicted that the BLA is constant with the input spectrum in memoryless nonlinear systems, three different spectrums are used, to compare it with the BLA of dynamic systems.

Figure 5.5 depicts the input spectrum of each test signal used. In the same figure, the output spectrum is also plotted, where the adjacent-channel distortion generated in the nonlinearity can be observed.

As can be seen in Figure 5.6 the BLA is unaffected by the input signal spectrum shape. This confirms the theoretical results previously obtained in Section 5.4, which state that the $H_L(\omega)$ of (5.7) is only dependent on the even order moments of the input signal and not on its shape.

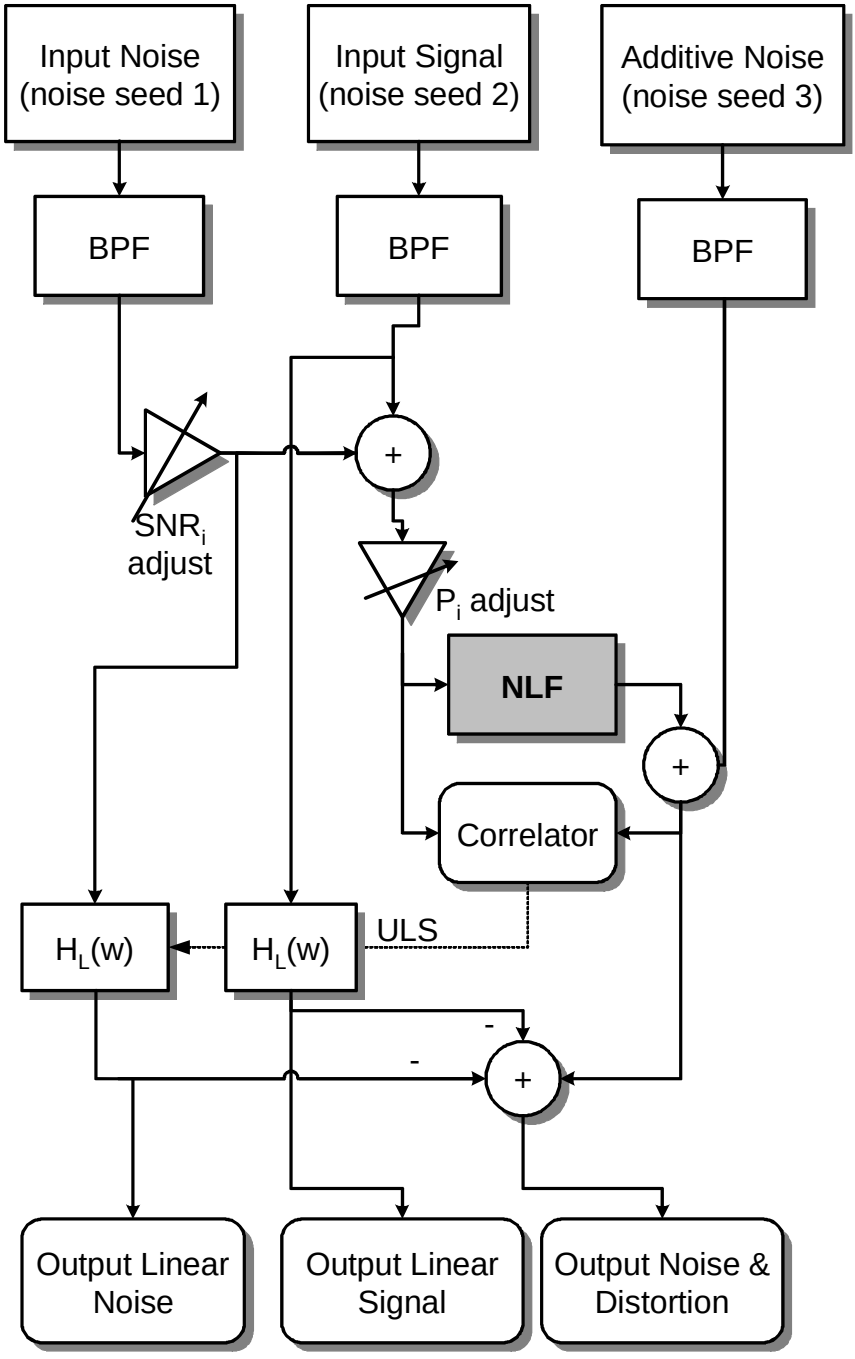


Figure 5.4 – Block Diagram of the simulator used to validate NDF, NLF stands for Nonlinear Function.

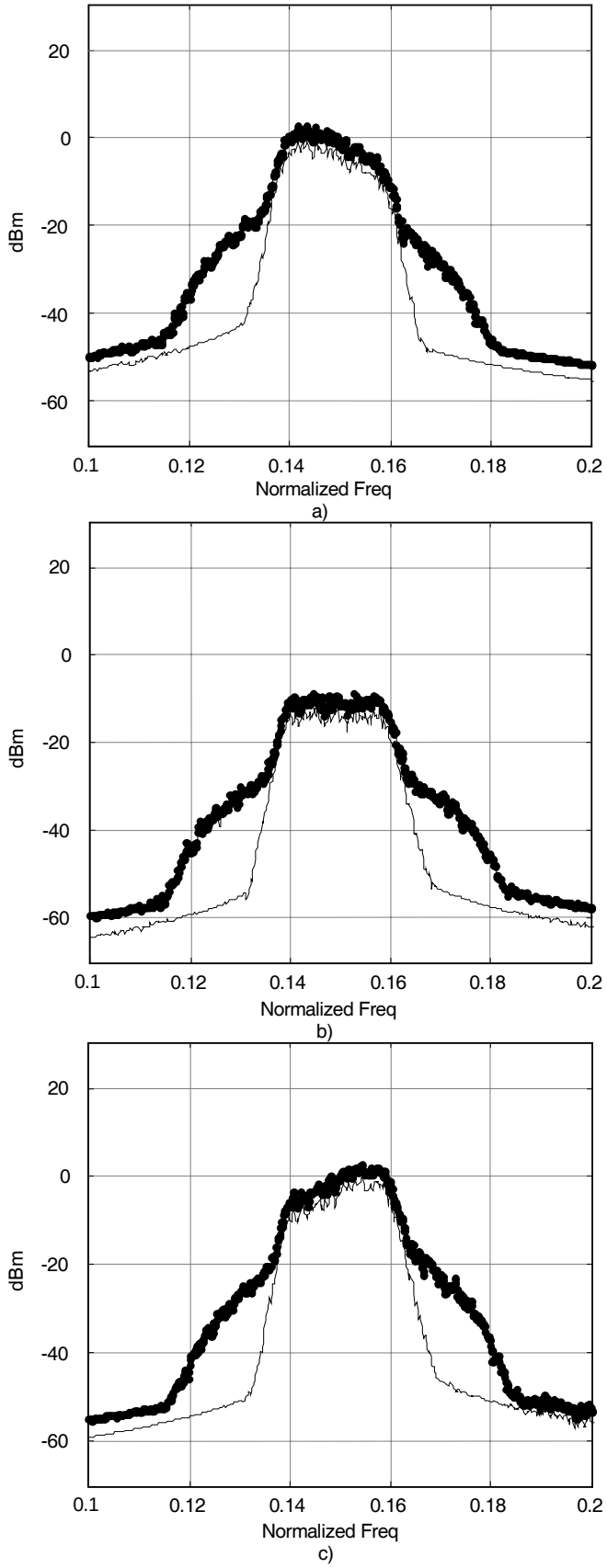


Figure 5.5 – The input spectrum of the test signals used for the BLA extraction (simple line) and Output Spectrum (dark line). a) Signal Spectrum 1. b) Signal Spectrum 2 c) Signal Spectrum 3

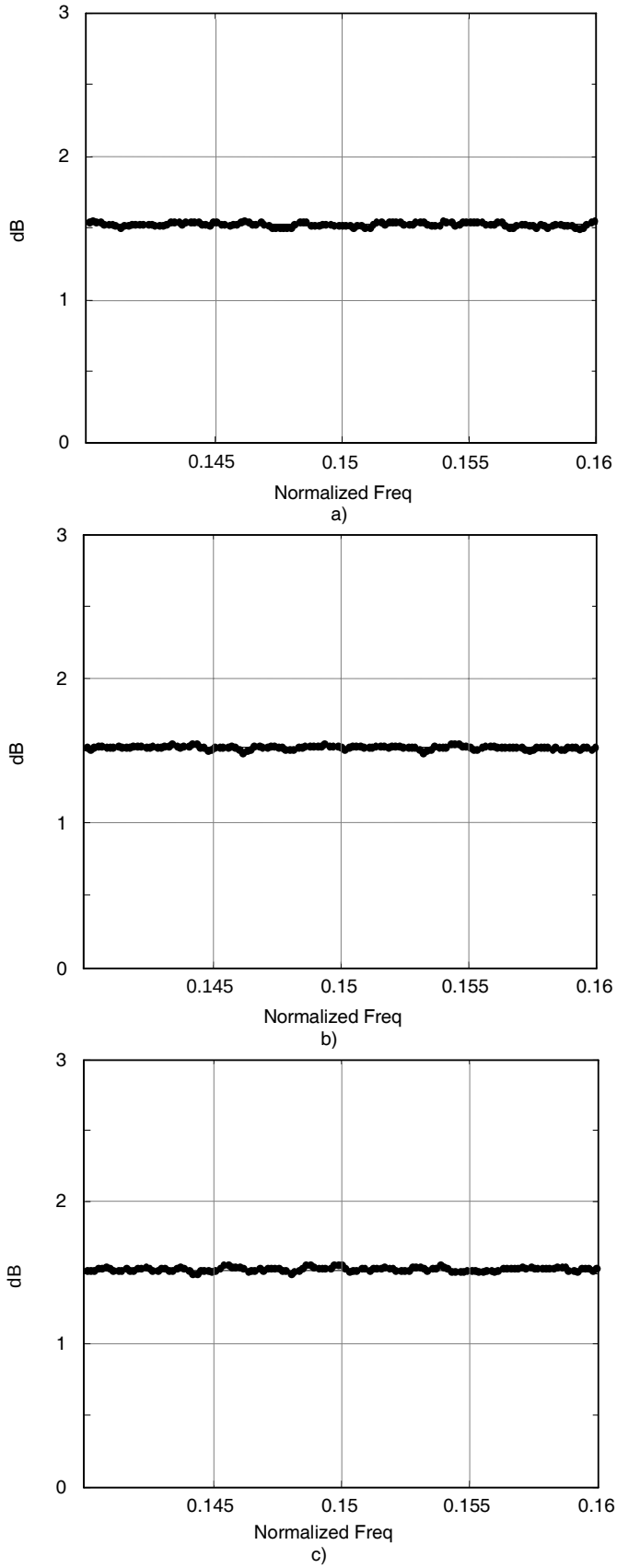
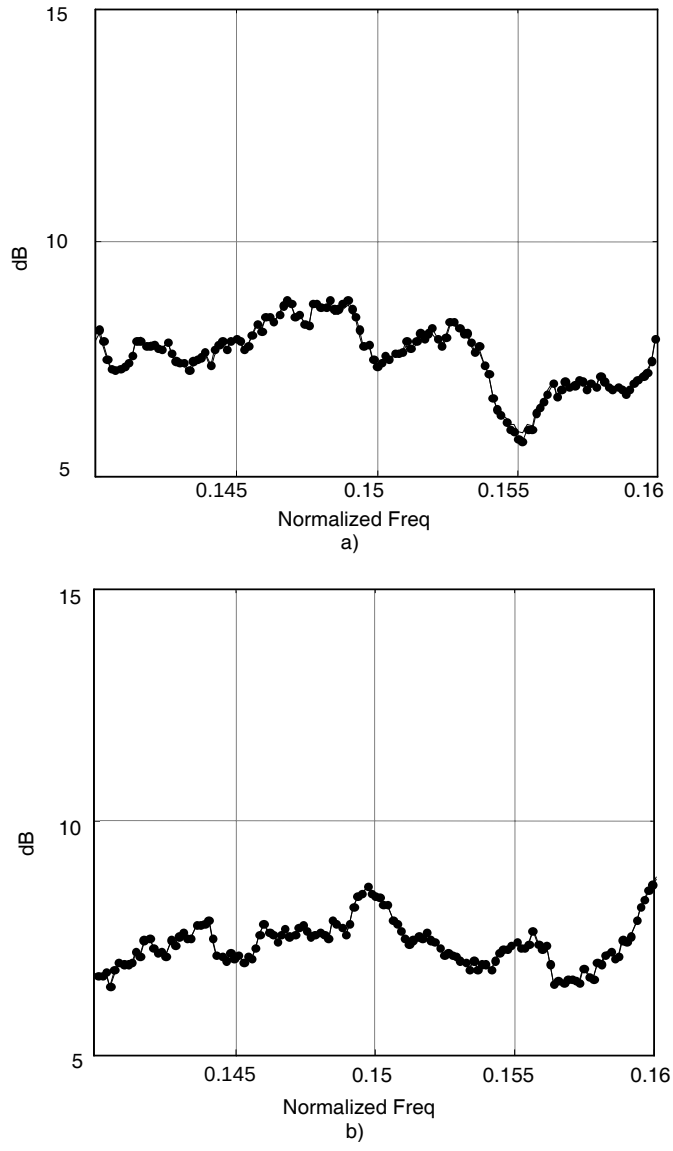


Figure 5.6 – In-band BLA: simulated (simple line); theoretical (dark line). a) Signal Spectrum 1. b) Signal Spectrum 2 c) Signal Spectrum 3

Although the BLA is an invariant to the input spectrum shape, the NDF varies with it, as shown in Figure 5.7. This variation is due to the fact that, in the frequency zones where the input signal spectral density is higher, the nonlinear distortion level increases at a faster rate than the output signal. Hence, the NDF must also present higher values.



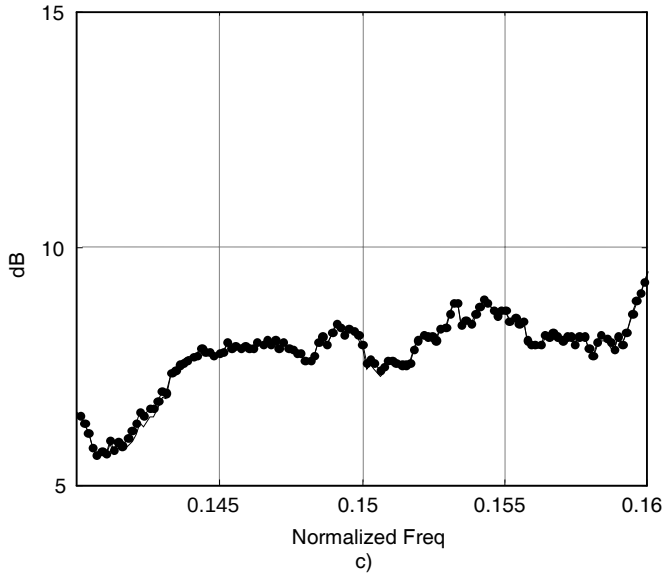


Figure 5.7 – In-band NDF: simulated (simple line), theoretical (dark line).a) Signal Spectrum 1. b) Signal Spectrum 2 c) Signal Spectrum 3.

5.6.2 NDF Calculation in a nonlinear system with memory.

In order to calculate the BLA for a nonlinear system with memory, a polynomial with α_1 and α_3 identical to the memoryless case was used, but the α_2 was increased to give more emphasis to the memory effect that is being studied [5.14, 5.15]. Note in expression (5.22) that, increasing the polynomial second degree coefficient, α_2, also varies the third order Volterra kernel [5.14]. A low frequency feedback filter $F(\omega)$ with frequency response shown in Figure 5.8 was introduced.

Note that this filter has a rejection ratio of more than 20 dB in the fundamental frequency zone and a steep roll off at the low frequencies between 0 and Bw/2 (Bw of the signals used were approximately 2% of the sampling frequency).

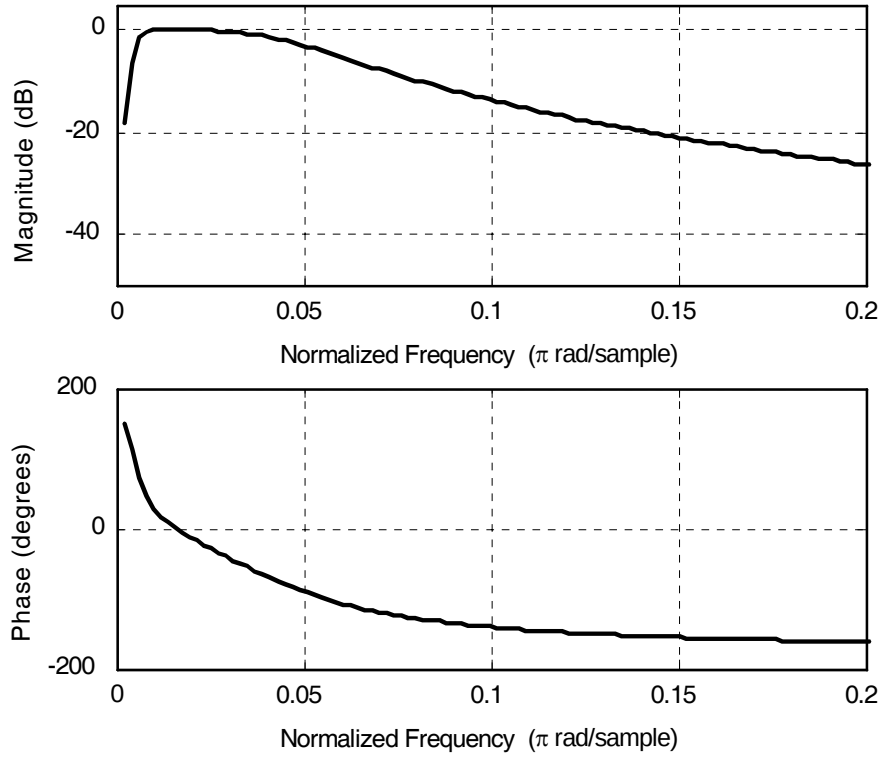


Figure 5.8 – Frequency response of the feedback filter F(ω) used.

Figure 5.9 shows the input and output spectra of the test signals used. The output has an adjacent distortion level higher than the one used when compared to the memoryless case (Figure 5.5). This is a consequence of the fact that, in this case, the strong second order coefficient also contributes to in-band distortion due to the feedback path, as seen in expression (5.22).

The valley shown in the BLA plot of Figure 5.10 b) is due to the high-pass characteristic of the feedback filter manifested between DC and 0.02, as $F(\omega)$ increases from the center of the band (DC) to the extremes (± Bw, the occupied signal bandwidth). This effect can only be noticed in this figure, because that case is the only resulting from a flat input spectrum. For the input spectra 1 and 3 [Figure 5.10 a) and c)], and due to the dynamic behavior of the feedback path, the BLA is affected simultaneously by the input spectrum and filter shapes, this way reducing the effect of the filter format.

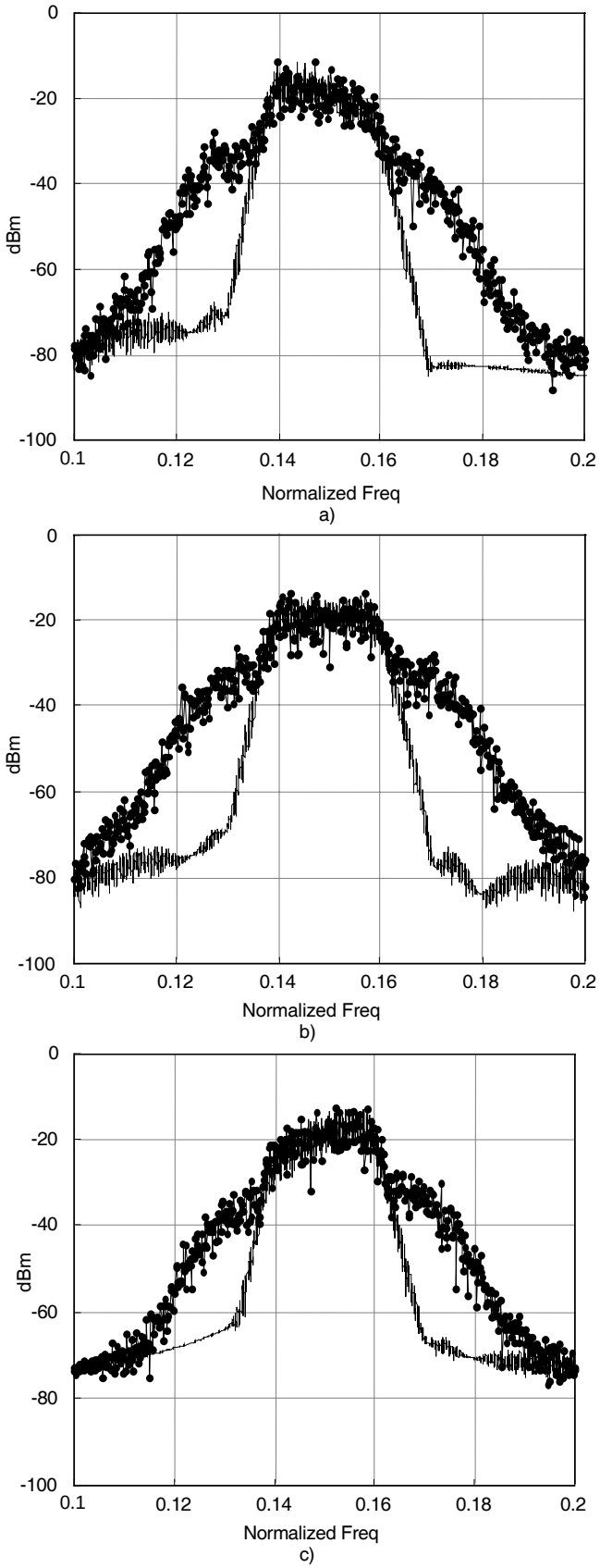


Figure 5.9 – The input spectrum of the test signals used for the BLA extraction (simple line) and Output Spectrum (dotted line). a) Signal Spectrum 1. b) Signal Spectrum 2 c) Signal Spectrum 3.

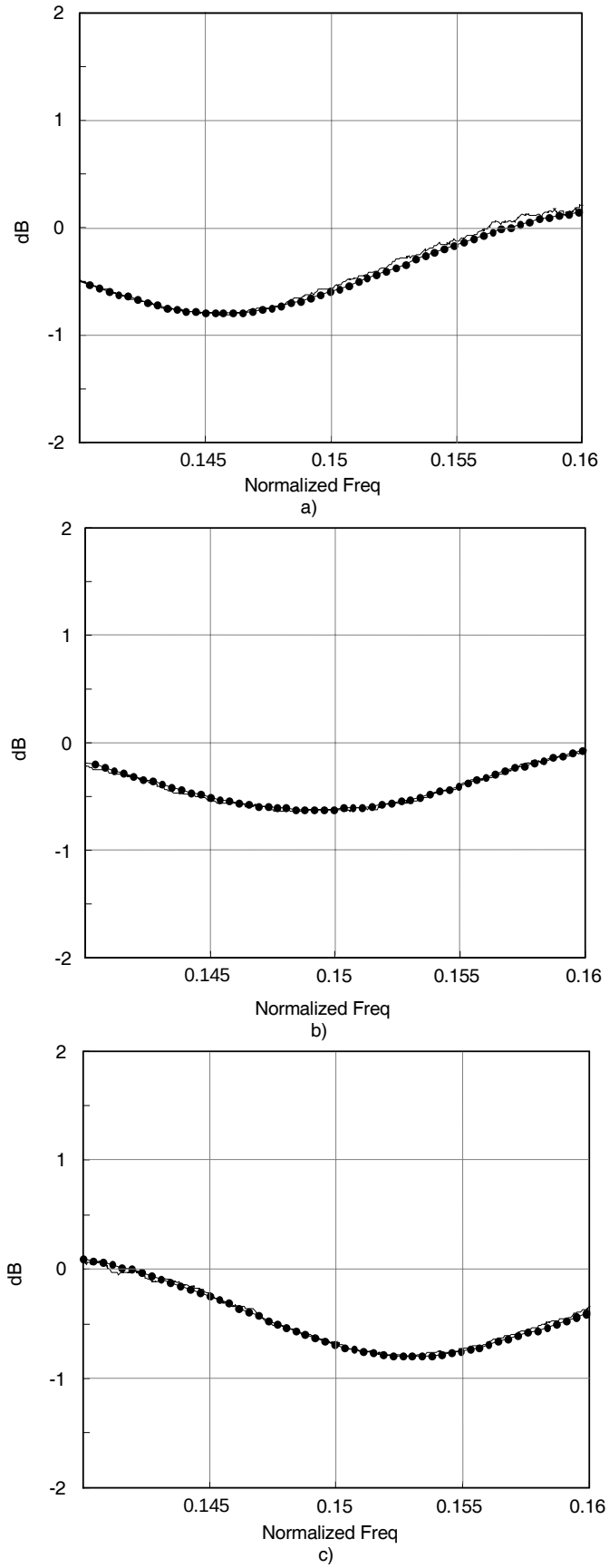


Figure 5.10 – In-band BLA: simulated (simple line); theoretical (dotted line) a) Signal Spectrum 1. b) Signal Spectrum 2 c) Signal Spectrum 3.

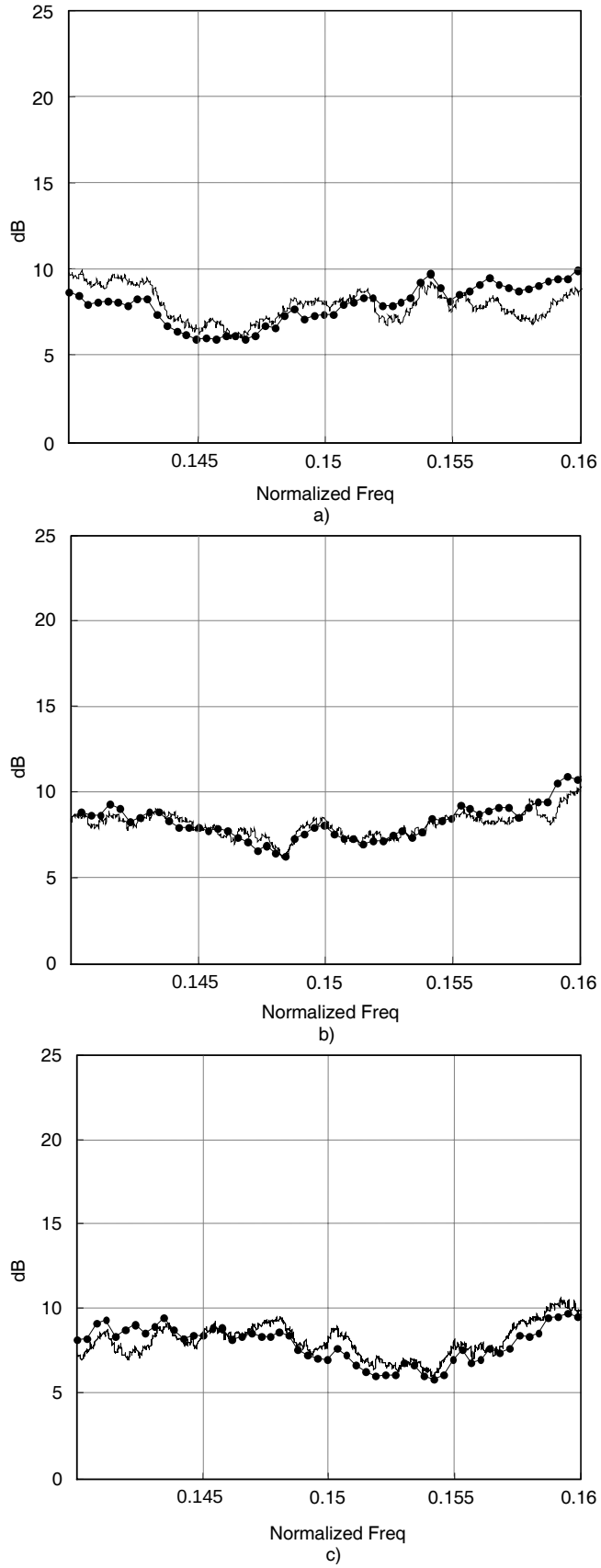


Figure 5.11 – In-band NDF: simulated (simple line); theoretical (dotted line) a) Signal Spectrum 1. b) Signal Spectrum 2 c) Signal Spectrum 3

The residual differences seen between theoretical and simulated results are due to the fact that the noise signal here in use is not of infinite length but a limited sequence whose realizations were averaged in frequency.

In Figure 5.11 the NDF is also presented and, as theoretically predicted, it also varies with the input spectrum.

5.7 Conclusions

In this chapter, to analytically characterize the NDF, the BLA was calculated for memoryless and dynamic systems. In the memoryless case it was shown that BLA is only dependent on the nonlinearity and on the input power. In dynamic systems, however, it also depends on the input spectrum shape.

The results obtained in this chapter for the BLA dependence on the input signal are a clear indication that particular attention must be paid to the input signal used to determine a nonlinear system model, since the system behavior (and thus the model) might be strongly dependent of the input excitation used.

Additionally in this chapter the misleading result presented on paper [5.2] was discussed and its main drawbacks pointed out. Nevertheless, the important conclusion obtained in [5.2], stating that the usual NF standard is affected by nonlinearities, is used to propose a new figure of merit, called NDF, that relates the input and output SINADs. Therefore, a NDF definition for nonlinear systems, but still consistent with the traditional linear NF, was advanced.

The excellent agreement between simulated and theoretical results gives us confidence to use this figure of merit in link budget designs.

5.8 References

[5.1] D. M. Pozar, *Microwave and RF Wireless Systems,* New York: John Wiley and Sons, 2001.

[5.2] A. Geens and Y. Rolain "Noise figure measurements on nonlinear devices," *IEEE Trans. on Instrumentation and Meas.*, vol. IM-50, no. 4, pp. 971-975, Aug. 2001.

[5.3] H. Ku, M. D. Mc Kinley, J. S. Kenney, "Quantifying memory effects in RF power amplifiers," *IEEE Trans. on Microwave Theory and Techniques,* vol: 50 Issue: 12, pp. 2843-2849, Dec. 2002.

[5.4] J. H. K. Vuolevi, T. Rahkonen, J. P. A. Manninen,, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques,* Vol. 49, pp. 1383 – 1389, Aug 2001

[5.5] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley & Sons, 1980

[5.6] R. Pintelon and J. Schoukens, *System Identification – A Frequency Domain Approach*, Piscataway: IEEE Press, 2001

[5.7] N. Blachman, "Bandpass nonlinearities," *IEEE Trans. Inform. Theory,* Apr. 1964, IT-10, (2), pp. 162-164.

[5.8] N. Blachman, "The effect of a large signal upon a small signal in a memoryless nonlinear bandpass amplifier," *IEEE Transactions on Communications*, Vol. 29 Issue: 1, pp. 72 -73, Jan 1981

[5.9] H. E. Rowe, "Memoryless nonlinearities with Gaussian inputs: elementary results," *The Bell System Tech. Journal*, vol. 71, no.7, pp 1519-1525, Sep. 1982.

[5.10] P. M. Lavrador, N. B. Carvalho and J. C. Pedro, "Noise and distortion figure – an extension of noise figure definition for nonlinear devices," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, U.S.A, 2003, pp. 2137-2140.

[5.11] - , *IEEE Standard Dictionary of Electrical and Electronics Terms*, Fourth Edition, November, 1988

[5.12] T. C. Hofner, "Defining and testing dynamic ADC parameters," *Microwaves & RF*, Nov. 2000

[5.13] J. C. Pedro and N. B. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*, Artech House, Inc., Aug. 2003

[5.14] J. C. Pedro, N. B. Carvalho and P.M. Lavrador, "Modeling nonlinear behavior of bandpass memoryless and dynamic systems," in *IEEE MTT-S Int. Microwave Symp. Digest*, Philadelphia, U.S.A, 2003, pp. 2133-2136.

[5.15] N. B. Carvalho and J. C. Pedro, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Trans. on Microwave Theory and Techniques,* vol. 50, Issue: 9, pp. 2090 – 2101, Sept. 2002.

6. Conclusions

Along with this thesis partial conclusions have been presented at the end of each chapter; so, now, is made a more general a summary of the work presented in this thesis.

In chapter 2 a description of PA main characteristics was made in order to introduce the main requirements for behavioural modelling. Then a brief overview of some behavioural model approaches and its approximation capabilities were discussed. Additionally, an introduction to Volterra series modelling approaches was presented including also a description of prior works on orthogonal Volterra series identification.

In Chapter 3 the derivation of the new orthogonal approach for Volterra series coefficients extraction was included. In this derivation the input signal for orthogonality was chosen and each model branch expression was derived. This derivation process included the general formula to build recursively any model order term. Also described in this chapter were the coefficient convergence problem and the conversion from the orthogonal model to the time domain Volterra series.

Different model validation simulation and measurement setups were presented in chapter 4. From the trivial memoryless nonlinear amplifier and linear filters up to real

microwave power amplifiers a vast set of validations were performed. The adopted model presents good approximation results in different situations but there is still margin for some improvements. Due to the adopted Volterra model topology the number of coefficients increases exponentially with the nonlinear order considered. This issue prevents the simultaneous use of a large number of memory span and nonlinear order.

Cross-correlation techniques, which enabled the separation of the output signal from the output noise, were presented in chapter 5. In this chapter the Best Linear Approximator was computed for systems modelled by a Volterra series. It was proved that, for a given system, the BLA is dependent on the input signal considered, and that both the input signal power and input signal statistics, influence the system's BLA. Furthermore, the definition of BLA enabled the proposal of a new Noise and Distortion Figure that allows the quantification of signal quality degradation on a system caused by additive noise and nonlinear distortion.

Five years ago, in the beginning of this work, when trying to understand the difference between correlated and non-correlated distortion, as well as what is the impact of each one on the signal quality degradation, I was very far from imagining all the problems that would need to be solved in order to reach the contributions presented in this thesis. Particularly, the orthogonal model formulation became an intricate problem that required a lot of dedicated and persistent work. Different approaches were tested more or less successfully, and a long discussion was carried on what should be an orthogonal model.

However, after these years of work, I think it is possible to truly state that the nonlinear distortion impacts are now slightly more understood, and that the behavioural modelling scientific field received some more contributions with well grounded formalism and guaranteed predictive capabilities.

6.1 Future Work

After this work there are still some open questions that might lead to some interesting work. The first one should be the extension of the modelling procedure validation, by comparing, for the same type of system, the now proposed technique with other conventional Volterra series extraction methods. This way the benefits of this

technique could be demonstrated comparatively with the existing ones, and the workload required to extract the coefficients by each of the methods could be compared.

It is known that one of the major disadvantages of the Volterra series is its large number of coefficients. Some interesting approaches have been proposed recently to prune the Volterra series [6.1-6.3]. It would be interesting to analyse from an orthogonally extracted series which are actually the negligible coefficients and after this to come up with a truncated version of a model and the corresponding extraction procedure that allows the orthogonal extraction of a truncated Volterra series.

The compromise between power efficiency and linearity in order to extend battery life and spectral efficiency has been driving the search for improved PA designs. A recent hot topic on this compromise, are the switched PAs that constitute a change in the paradigm of designing a PA. It is already known [6.4] that the distortion mechanisms for this type of amplifier are quite different from the ones of the "conventional" PAs. This way the development of new behavioural models able to mimic the nonlinear response of switched PA is a major question waiting for an answer.

The continue developments of integrated devices are leading to higher levels of integration and system complexity that create large challenges for chip and system design. The verification/simulation of these mixed-signal high integrated devices is an unacceptably large computational task, which requires the replacement of circuit-level models by accurate, all-purpose behavioural models. In this scenario, the application of the knowledge gathered in this thesis to the modelling of complete integrated systems or subsystems should also be evaluated.

6.2 References

[6.1] A. Zhu, T.J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned volterra series", *IEEE Microwave and Wireless Components Letters*, vol. 14,Issue 12, pp. 563 – 565, Dec. 2004.

[6.2] A. Zhu, J. Dooley, T.J. Brazil, "Simplified volterra series based behavioral modeling of RF power amplifiers using deviation-reduction," in *IEEE MTT-S International Microwave Symposium Digest*, June 2006, pp. 1113 – 1116.

[6.3] A. Zhu, J.C. Pedro, T.J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. On Microwave Theory and Techniques*, vol. 54, Issue 12, Part 2, pp.4323 – 4332, Dec. 2006.

[6.4] J.C. Pedro, J.A. Garcia, P.M. Cabral, "Nonlinear distortion analysis of polar transmitters", *accepted for publication at the IEEE MTT-S International Microwave Symposium*, Honolulu 2007.