Behavioral Modeling of Unmatched Nonlinear Devices Driven with Modulated Signal Stimuli using Volterra Series

by

Marwen Ben Rejeb

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

©Marwen Ben Rejeb 2014

AUTHOR’S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Marwen Ben Rejeb

Abstract

The accurate simulation of nonlinear radio frequency (RF) circuits under unmatched impedance conditions depends heavily on the device model used. Recently, measurement-based models under unmatched conditions were proposed in the literature, such as X-parameters and the Cardiff model, with acceptable device modeling accuracy under continuous-wave (CW) and multitone excitations. However, these schemes drop in accuracy when driven with wideband modulated signals since the pseudo-static property of the CW stimulus fails to capture the nonlinear device dynamics. Being a general framework to model nonlinear dynamic systems with fading memory, the Volterra series was successfully applied to model nonlinear RF devices under matched conditions. However, the few reported attempts to generalize the Volterra series to unmatched device scenarios oversimplify the formulation due to the underlying complexity burden.

This thesis presents a novel low-complexity envelope domain multivariate model (i.e., dual-input dual-output or DIDO Volterra model), tailored to unmatched devices. The proposed model offers a theoretical framework to model the nonlinear device's behaviour in any two-port network around the fundamental frequency. As the proposed methodology does not include any assumptions about the device under test (DUT), the resulting model can be applied to either RF transistors or power amplifiers (PAs). The detailed derivation steps, as well as the resulting mathematical formulations, are presented and discussed. For a fair assessment of the model's performance, a simulation procedure was developed where a PA under varying load impedances was modeled.

Any modeling exercise is tightly coupled to both the device characterization and measurement procedures. Indeed, the successful application of a behavioral model is contingent on a precise extraction of the nonlinear measurements. In the case of the DIDO Volterra model, the use of modulated signals as stimulus further complicates the measurement procedure and a more careful extraction procedure should be implemented. The platform should be able to correctly measure the travelling waves' envelopes at the device ports while controlling the source/load impedances. Hence, the second part of this thesis outlines progress towards the realization of such a platform and focuses on explaining the associated challenges. A thorough discussion of the platform architecture and the underlying building blocks is presented and the related choices justified. Furthermore, additional calibration routines under wideband signals are described.

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to my supervisor, Dr. Slim Boumaiza, for his guidance and insights throughout my master. His advice, professionalism and encouragement helped me to improve and to always challenge myself. Also, I would like to thank Dr. Lan Wei and Dr. Shreyas Sundaram for reading my thesis and giving their valuable comments.

I cannot describe how grateful and blessed I feel to have such wonderful friends. I thank them for their support, love and laughs. To my parents and family, the thought of making you proud is what wakes me every morning. Thank you for being part of my life.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACPR	Adjacent Channel Power Ratio
ADC	Analog to Digital Converter
ADS	Advanced Design System
ANN	Artificial Neural network
AWG	Arbitrary Waveform Generator
CAD	Computer Aided Design
CW	Continuous Wave
DAC	Digital to Analog Converter
DDR	Dynamic Deviation Reduction-Based
DIDO	Dual Inputs Dual Outputs
DUT	Device under Test
HSP	Harmonic Superposition Principle
IF	Intermediate Frequency
IMD	Inter-Modulation Product
LNA	Low Noise Amplifier
LO	Local Oscillator
LSE	Least Square Error
LSNA	Large Signal Network Analyzer
LSOP	Large Signal Operating Point
MIMO	Multiple Input Multiple Output
NMSE	Nominal Mean Square Error
NVNA	Nonlinear Vector Network Analyzer
PA	Power Amplifier
PAPR	Peak to Average Power Ratio
RF	Radio Frequency
SFDR	Spurious Free Dynamic Range
SISO	Single Input Single Output
SOLT	Short-Open-Load-Thru
TRL	Thru-Reflection-Line
VNA	Vector Network Analyzer

Chapter 1

Introduction

1.1 Motivation

The number of connected devices has seen exponential growth during the last decade. Smartphones and tablets, currently estimated at around 1.5 billion with a steady increase in this number projected, are pushing towards more data speed and more network capacity. In addition to the voice services, smart device users increasingly rely on internet-based services (email, streaming, social networks…) and expect reliable and high data rate throughout the network coverage. To satisfy this demand, for each generation of mobile telecommunication technology, the RF front-end specifications are increasingly stringent in terms of operating frequency, peak to average power ratio (PAPR) and signal bandwidth. As an illustration, the nominal bandwidth of GSM standard (2^{nd} generation) is 200 KHz while LTE-Advanced standard (4^{th} generation) supports up to 5 carriers and up to 100 MHz bandwidth, which enables download speeds of up to 1 Gbit/s for low mobility communication scenarios.

These tight specifications dictate strict linearity and efficiency requirements for RF transceivers and further complicate the design flow of any microwave circuit. In fact, achieving higher efficiency requires the device be driven into the nonlinear region which creates a dilemma for the designer who has to make a tradeoff. Therefore, the linear and narrowband approach for designing microwave circuits which is based on an iterative process of designing, testing, then tuning is simply no longer applicable. The solution resides in the utilization of computer-aided design (CAD) tools, now mandatory for simulating and optimizing the design to avoid the lengthy *measure & test* approach and reducing the overall time-to-market. Despite the advances made in recent CAD simulators, the agreement between simulation and measurement results has been an issue for designers. The roadblock to the accuracy of these simulators lays predominantly in the precision of the used nonlinear device model, especially the transistor model. Therefore, any effort to improve the simulators performance goes through the process of developing accurate nonlinear device models.

Nonlinear device models could be grouped under three categories: phenomenological or physical models, compact models and behavioral models. Phenomenological models are inspired from the physics of the devices while compact models provide a circuit representation that mimics the behavior of the device. Behavioral models is based on black-box mathematical approximations that map the outputs of the device to its inputs independently of their generation mechanism. Behavioral models have been widely adopted as they are technology-independent and have significantly faster system-level simulation. In the context of unmatched devices, like it is the case for transistors, measurement-based behavioral models are generally extracted under CW excitations based on different operation conditions (device dimensions, DC biasing, carrier frequency, input power) as well as source/load terminations.

It is important to be aware that successful behavioral models are highly dependent on the availability of advanced characterization platforms. Nonlinear measurement setups that can characterize nonlinear RF circuits in terms of power gain compression, efficiency, output power, intermodulation products under different source/impedances, input power stimulus and bias conditions have been developed following several configurations. Hence, advances in the modeling generally follow breakthroughs in the characterization techniques and specialized test equipment. Traditional measurement systems cannot provide voltages and currents information at the ports of the DUT. However, the recently available nonlinear vector equipment can perform

unrationed vector measurements, i.e. amplitude and phase, while preserving the relative phase between fundamental frequency and its harmonics [1]. Capturing all spectrum components and ensuring their relative phase coherency enables the reconstruction of time domain signals and provides the designer with valuable information. The emergence of nonlinear vector equipment greatly boosted the development of behaviour models for nonlinear devices.

Despite the progress made in the characterization of nonlinear devices, the discrepancy between the CW or multitone and wide-band modulated signal stimulus creates a major handicap and further complicates the modeling process. Indeed, the conditions under which the DUT is excited for model parameters extraction differs from the conditions of its deployment resulting in unpredicted results. Therefore, there is a need for developing a novel characterization platform along with behavioral models, both using modulated signal, to help overcome this limitation.

1.2 Problem Statement

This thesis targets in particular two nonlinear devices: transistors and RF PAs. However, the concept developed can be applied to any RF two-port device. Transistor models enable the RF circuit designer to perform accurate simulations. Furthermore, accounting for the mismatch in RF power amplifiers should provide a tool for system level simulations to properly conduct multi-stage RF power amplifier design and to predict the overall behavior of a RF power amplifier cascade with an another nonlinear device. For example, it should be possible to predict the effect of a change in the antenna's impedance that follows the RF power amplifier.

As stated previously, behavioral models for unmatched nonlinear devices are generally extracted under CW harmonic excitations. CW stimulus static nature includes only for the quasi -static response of DUT and excludes the device dynamics, which compromises the model accuracy when used under realistic modulated signal excitations. With wideband signals, the distortions attributed to memory effects increases and cannot be neglected anymore. In the case of a transistor, memory effects are the combination of different physical phenomena such as trapping effects, biasing network and any matching network associated with the transistor. They are generally classified as short-term memory effects or long-term memory effects depending on their time constant that ranges from nanoseconds to seconds. Due to memory effects, the nonlinear device's response is no longer solely dependent on the instantaneous voltages and currents at the port of the device but also depends on the past instances of the voltages and currents. Moreover, the thermal behaviour differs between CW excitation and modulated signal excitation and capturing the self-heating mechanism becomes a mandatory task. In order to solve this issue, a novel model was developed, to account for the dynamics that the DUT exhibits under modulated signals stimulus using model formulation and extraction around modulated signals.

Rather than starting from a static formulation and trying to augment the model in order to account for memory effects and thermal behaviour which can be a complicated exercise, this work uses Volterra series as general framework that inherently captures nonlinear devices' dynamics with fading memory and aims instead to reduce its complexity to a practical degree. Indeed, in the context of matched environment, like RF PAs, behavioral models are often based on polynomial functions approximation (derivatives of the Volterra Series) or artificial neural network (ANN) driven with realistic modulated-signals. Therefore, this thesis focuses on the extension of the Volterra series formalism from the single-input single-output (SISO) context for matched nonlinear devices to the context of dual-inputs dual-outputs (DIDO) unmatched nonlinear devices.

The motivation behind this expansion is to take advantage of the success of Volterra series, proven in the context of matched environment especially its ability to model memory effects, and apply it to unmatched environment. The extension takes particular care controlling the complexity

of the model in order to enable its practical deployment in CAD environment given that Volterra series complexity is already an issue in the SISO formulation. The new model, extracted under modulated signal, should yield more accurate system level simulation with reasonable complexity.

The DIDO Volterra model uses a modulated signal as stimulus imposing the development of nonlinear characterization setup that uses modulated signal stimuli. The development of this setup is not straight forward and involves the use of advanced measurement equipment as well as sophisticated calibration and validation routines since errors in measurements can significantly degrade the results of the model. Hence, in addition to the DIDO Volterra model, this thesis compiles guidelines, calibration procedures, challenges investigation and initial steps towards the development of such platform.

1.3 Thesis Organization

The thesis is organized in five chapters. A brief summary of the contents of each chapter is presented below:

- *Chapter 1:* This chapter represents an introduction that motivates the work presented and gives an insight on the challenges that had to be overcome as well as the expected results. It clarifies the problem at hand and its relevance. It also focuses on the novelty in the proposed solution.
- *Chapter 2:* The second chapter summarizes state-of-the-art nonlinear measurement setups as well as nonlinear device behavioral modeling approaches used in the literature. First, the measurement setups are compiled and compared in terms of their capabilities, measurement time, harmonics, etc. Secondly, the modeling approaches are classified in two categories: The first includes, the models used under the describing functions framework for unmatched environment, where X-parameters model and Cardiff model are taken as examples, whereas the second contains the models based on Volterra series framework for matched environment. In the last section of this chapter, a comparison between the two modeling frameworks is conducted.
- *Chapter 3:* The third chapter focuses on the newly developed DIDO Volterra model. The steps to establishing the model formulation and derivation from the SISO Volterra model are explained. Also, the reduction in the complexity of the model is illustrated. Finally, the phases of extraction procedure are detailed and the simulation results using a CAD commercial simulator, Agilent Advanced Design System (ADS), in terms of nominal mean square error (NMSE) and frequency spectrum for validation purposes are presented.
- *Chapter 4:* Chapter four describes the measurement setup for extracting DIDO Volterra model using modulated signals. The challenges and guidelines for implementing the setup are discussed. Calibration routines and measurement procedures are illustrated. Experimental results are also provided using in-lab equipment.
- *Chapter 5:* The final chapter includes a summary of the important aspects discussed in this work concerning both modeling and characterization and explains the future steps that could be performed to improve the current results.

Chapter 2

Nonlinear Device Characterization and Behavioral Modeling Approaches

2.1 Nonlinear Characterization and Behavioral Modeling Overview

Extracting a phenomenological transistor model or a compact model has always been a complex and time-consuming task. The difficulty is worsened by the fact that these types of models are technology-dependent and therefore need to be re-extracted with each advance in the semi-conductor technology. Moreover, when implemented within a CAD environment to perform system level simulation, they need a long and sometimes unpractical simulation time. For all the previous reasons, behavioral modeling is an alternative solution. In fact, behavioral modeling is a technology-independent, black-box framework that significantly simplifies the modeling procedure of nonlinear devices. If extracted under the same conditions as the application conditions, behavioral models yield accurate performances since they are based on the interpolation and extrapolation capabilities of the global mathematical functions used.

Different possible classifications were suggested in literature for the behavioral models. The classification adopted in this thesis is based on the matching conditions of the DUT and results in two main categories:

- *Modeling of Nonlinear Unmatched Devices*

The modeling of nonlinear unmatched devices has to take into account the effect of the source/load impedances in order to predict the behavior of the DUT under different terminations. If the DUT was linear, extracting S-parameters under a certain source/load impedances would be sufficient to predict the behavior of the DUT under any other source/load impedance. Unfortunately, it is not the case for nonlinear devices. The typical example is when the DUT is a transistor. In that case, the performance of the transistor (gain, output power, efficiency, noise …) are a function of the source/load impedance and finding the right impedance is an important step of the RF nonlinear circuits design. The availability of a model for unmatched environment that predict accurately the behavior of the transistor avoids the long series of load-pull measurements.

- *Modeling of Nonlinear Matched Devices*

Models for matched nonlinear devices include system level components of RF transceivers like PA, low noise amplifier (LNA), mixers …. In the case of RF PA, they are used to derive pre-distorter and post-distorter in order to linearize the RF PA [2]. Figure 2.1 shows a representation of two-port network using power waves where a_1 and a_2 are the incident power-waves while b_1 and b_2 are the emerging power-waves. In the context of nonlinear matched devices, a_2 is equal to zero.

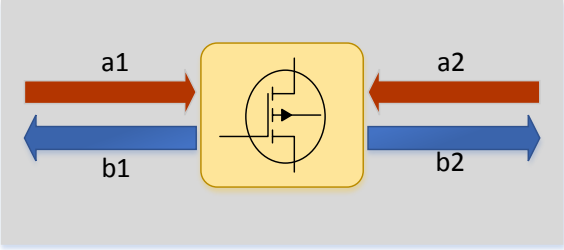


Figure 2.1 - Two-Port Device Representation

Indeed, b_1 results from a mismatch at the input of the device while a_2 is the generally a portion of the b_2 reflected back to the device because of a mismatch at the load side.

In the case of matched nonlinear devices, the measurement procedure is not complicated. A signal generator produces the modulated signal at the fundamental frequency. The signal goes through a driver if it needs to be amplified. Drivers are linear power amplifier that should not introduce any additional distortions to the signal. Afterwards, the signal is fed to the DUT and then to the receiver after the appropriate attenuation is applied. The whole measurement is conducted at the instruments impedance (generally 50 Ohm) at all frequencies. The challenges of such measurement are related to the wide bandwidth acquisition, dynamic range and high fundamental frequencies. The measurement setup is illustrated in Figure 2.2.

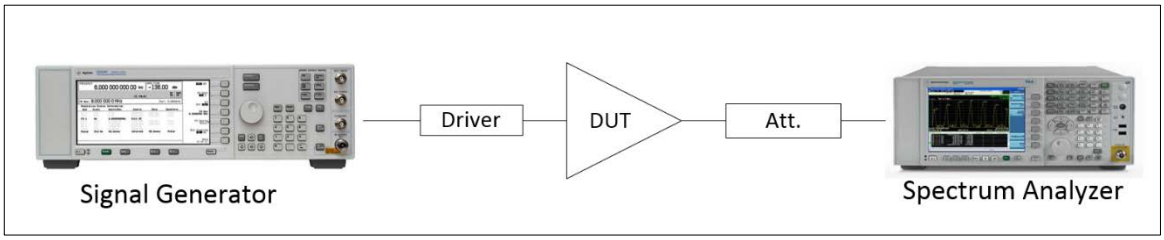


Figure 2.2 - Measurement Setup for Matched Nonlinear Devices

However, in the case of unmatched nonlinear devices, the measurement setup are much more complicated because the measurement system should be able to present different and controlled source/load impedances at the port of the device. Moreover, the measurement system is expected to measure the current and voltage or equivalently the power waves (A-waves and B-waves) at the port of the DUT. Hence, to develop behavioural models and even compact models device that are not available from the manufacturers, to represent behavior under unmatched conditions, diverse nonlinear characterization systems have been developed during the last decades. Due to the importance of these test benches, a large number of publications exists where each one of these publications tackles a particular aspect in the measurement procedures or exploits the advances in measurement equipment.

This chapter is structured in two parts. The first part gives an overview of nonlinear unmatched characterization systems structure followed by a state-of-art review of available characterization setups. The second part deals with the modeling aspect. It details the describing functions framework for modeling nonlinear unmatched devices. Then, it presents Volterra series framework for nonlinear matched devices. Finally, it concludes by comparing the diverse aspects of these two frameworks which will serve as an introduction and motivation to the DIDO Volterra model formulation.

2.2 Nonlinear Characterization Setups

2.1.1 Nonlinear Characterization Setups Structure

The variations of the nonlinear characterization systems, as described in the literature, are classified under three major categories, namely the type of stimulus used to excite the DUT, the measurement equipment used to measure the RF signals and the waveform engineering procedure used to control the source/load terminations at the DUT ports. Figure 2.3 summarizes the three parts of any nonlinear characterization setup and provides some examples under each category.

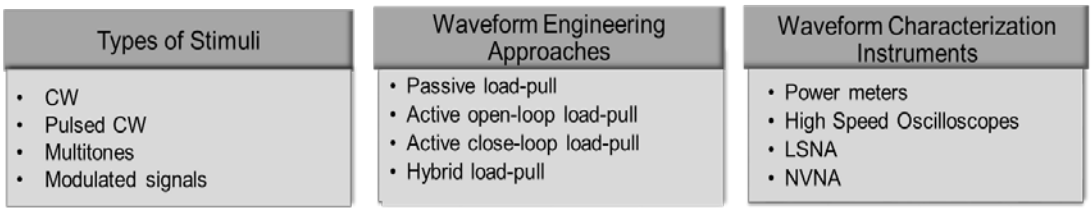


Figure 2.3 - Three Parts of Nonlinear Characterization Setups

2.1.1.1 Waveform Generation

Traditionally CW signals were used to characterize nonlinear devices. However, Pulsed CW, multitone and recently modulated signals where used as well in few works.

Using CW signals, performances like power gain and efficiency are measured against input power and ports terminations among others. However, CW stimulus cannot distinguish between the static nonlinearity of the DUT and its thermal behavior. Indeed, when driving a transistor under CW stimulus, it is subject to self-heating and its electrical characteristic changes significantly. To overcome this limitation, multiple characterization setups are developed using pulsed CW measurements. In a pulsed measurement the DC bias or/and the RF signal are pulsed with a certain pulse width where pulse width is the ratio between the *on* time to the *off* time during the pulse period. When the pulse period is narrow enough, isothermal behavior is achieved. Figure 2.4 shows the IV curves of an LDMOS transistor with pulsed signals and CW signals and how the behavior differs.

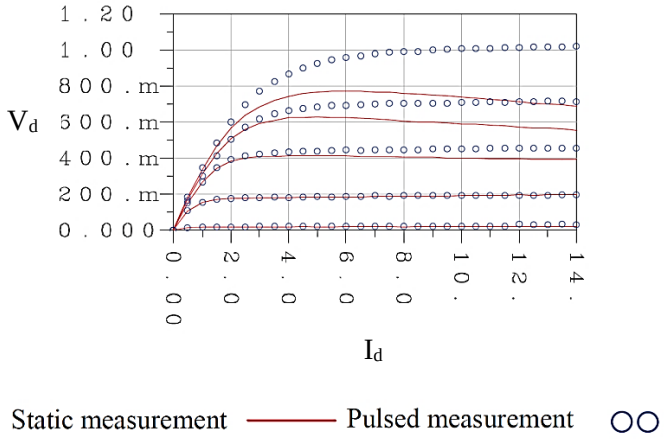


Figure 2.4 - Static IV vs. Pulsed IV measurement for a LDMOS Transistor [3]

Another benefit of pulsed measurements is that it alleviates the power handling requirement of the measurement system when characterizing high level devices as high power CW can damage the system or damage the device under test due to self-heating.

When concerned with the evaluation of the linearity of the DUT, a typical test consists in using multitones excitations and measuring the resulting intermodulation products. For example, if a two-tone test is performed on a nonlinear device at frequency f_1 and f_2, the nonlinearity will create frequency components, called intermodulation products (IMD), at $nf_1 + mf_2$ of order $n + m$. The relative power of the 3rd order intermodulation products, $2f_1 + f_2$ and $2f_2 + f_1$, to the fundamental tones power (IMD_3) is often used as metric to quantify the degree the DUT nonlinearity (see Figure 2.5). The higher the IMD_3, the stronger the nonlinearity.

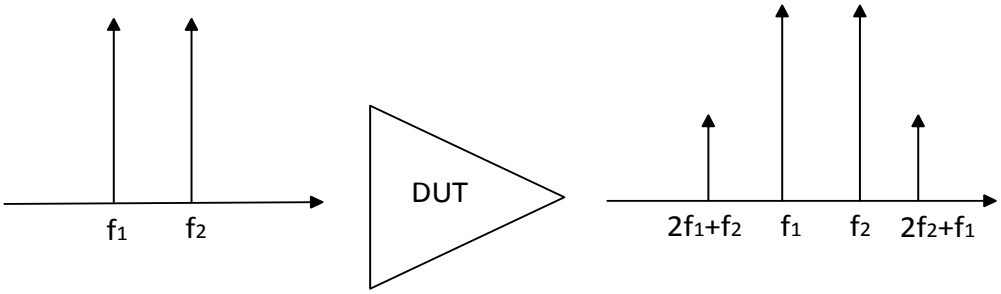


Figure 2.5 - DUT 3rd Order Intermodulation

As discussed before, CW characterization is limited to the quasi-static electrical behaviour and ignores the device dynamics. For this reason, more and more work focuses on multitone and modulated stimulus to characterize the memory effects and ensure more realistic performance to the extracted models when used with wideband modulated signals. When generating CW signals, multiple signal generators are combined and phase synchronized to produce the desired signals at fundamental and harmonics. However, for wideband signals (Multitone and modulated signals), arbitrary waveform generator (AWGs), which are basically high-speed digital to analog convertors (DACs), are used to create the baseband signal that is up-converted to the fundamental frequency.

2.1.1.2 Waveform Measurement Instruments

Waveform measurement instruments measure the currents and voltages at the DUT reference plans and provide frequency domain or/and time domain information under various operation conditions. Waveform measurement instruments are classified in three main categories:

- *High-speed oscilloscopes:* Scopes measure the waveform in time domain. They are limited by the achievable sampling speed. In order to satisfy Nyquist–Shannon sampling theorem, the sampling frequency of the oscilloscope has to be more than twice the highest harmonic measured which can easily reach few GHz. Since they capture the time domain signal, oscilloscopes measure all the spectrum components but they suffer from a low dynamic range which is usually around 60 dB and sampling clock jitter issues at very high frequencies. (see Figure 2.6)

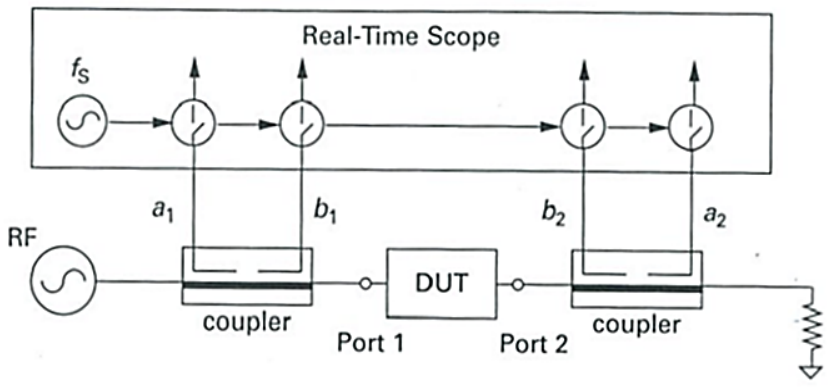


Figure 2.6 – High-speed Scope Based Characterization Setup [1]

Large Signal Network Analyzers (LSNA): LSNA measures a low-frequency alias of the high frequency signal's harmonics through the process of sub-sampling. The sampling frequency is generally around 20 MHz which is significantly lower than oscilloscopes. LSNA can only measure band-limited signals otherwise they suffer from aliasing. They can have a moderate

dynamic range but it still lower than the nonlinear vector network analyzer (NVNA). They also suffer from limited phase accuracy. (see Figure 2.7)

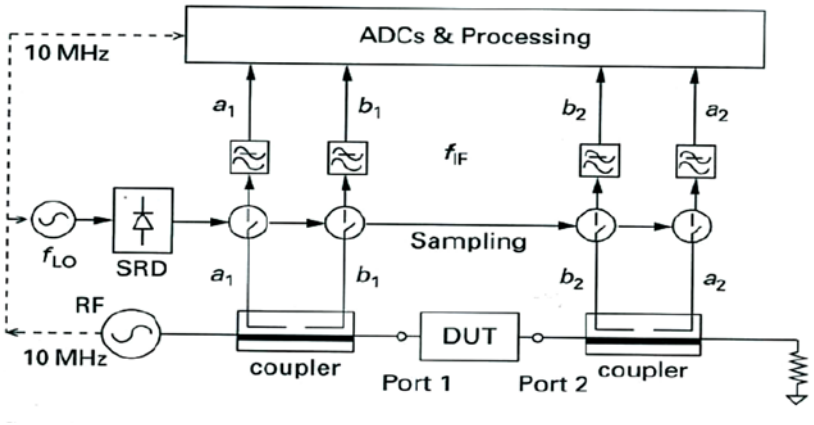


Figure 2.7– LSNA Based Characterization Setup [1]

- *Nonlinear Vector Network Analyzers (NVNA):* Based on VNA technology, the NVNA follows a heterodyne architecture to measure the magnitude and phase of the signal's frequency components one at a time in a phase coherent manner. Agilent NVNA for example use an external comb generator to ensure the phase coherency between fundamental and harmonics. NVNA has the highest dynamic range but the successive procedure of spectrum measurement can lead to long measurement time. (see Figure 2.8)

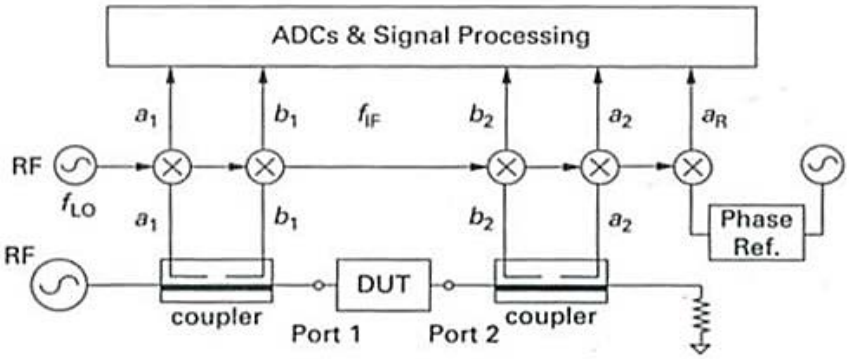


Figure 2.8 – NVNA Based Characterization Setup [1]

Each type of these instruments is used extensively for CW, pulsed CW and even multi-tone applications, but face major issues when the signal bandwidth is increased to the range of 100 MHz and beyond.

2.1.1.3 Waveform Engineering Approaches

Waveform engineering approaches implies the ability to synthesize specific voltages and currents waveforms at all ports of the DUT. For a two-port network, such as transistors, waveform engineering at the load side is called "*load-pull*" and at the source side is called "*source-pull*". Different techniques for source/load pull exist and they are classified in three categories:

- *Passive load-pull:* Historically passive load-pull are achieved using around tuners which are narrowband mechanical equipment composed of multiple stubs and stub lines that can be set to present a user-defined impedance at the ports of the DUT. Its simplicity of use is the reason behind its popularity within the industry. However, these tuners are slow to move from impedance to another and limited in their ability to provide high reflection coefficients

because of the losses between the tuners and DUT. Most of the time, the process includes a lengthy calibration phase. Multi-harmonic tuners are employed to control the reflection coefficients at a certain fundamental frequency and its harmonic frequencies. Figure 2.9 shows an example of passive tuners.

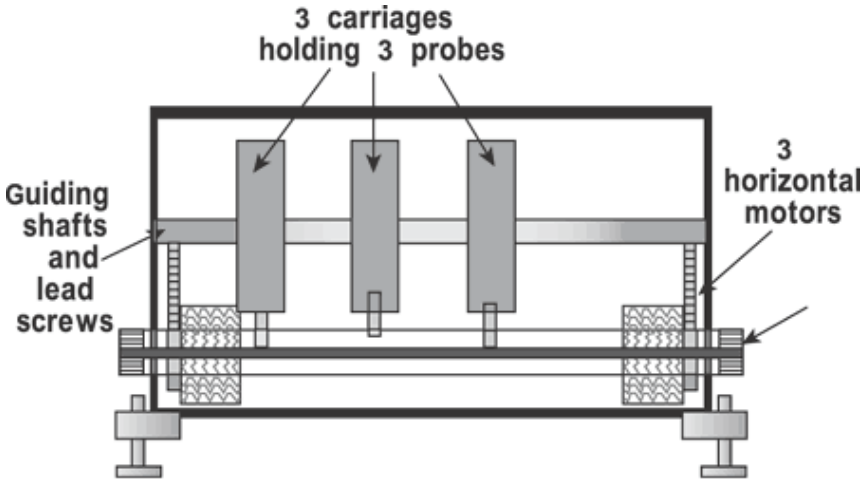
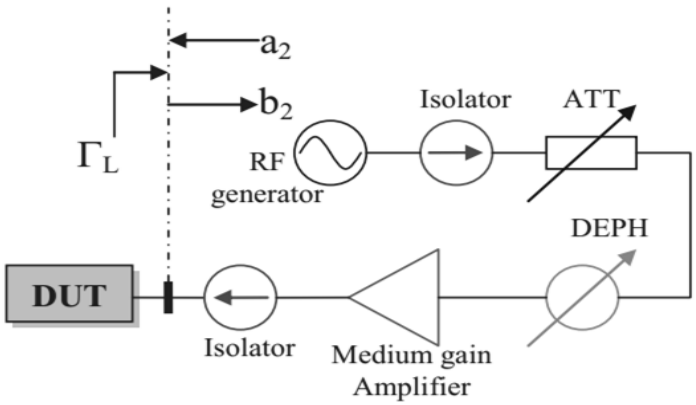
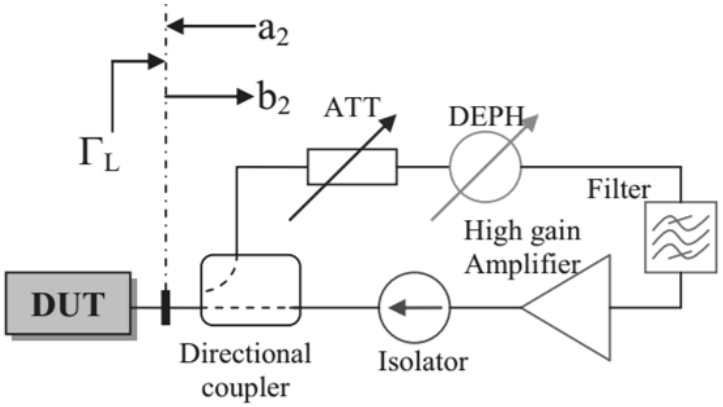


Figure 2.9 - Passive Tuner Diagram [4]

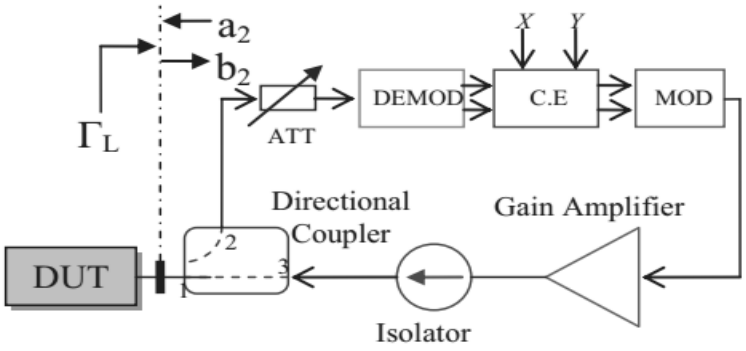
- *Active load-pull:* In order to overcome the limitations of passive tuners, especially the ability to cover the whole Smith chart, active load-pull has been gaining attention in recent years. In active load-pull, the signal at the output of the DUT is absorbed into a load or partially redirected to the DUT and another signal is injected to emulate a certain impedance. Active load-pull have been deployed using three different configurations: open-loop configuration and closed-loop configuration and envelope configuration (see Figure 2.10).



(a) Open-loop



(b) Close-loop



(c) Envelope load pull

Figure 2.10 - Active load Pull Configurations: (a) Open-loop, (b) Close-loop, (c) Envelope load pull [5]

In the open-loop active load-pull, the output power wave from the DUT is fully absorbed and a different signal is independently injected while in the closed-loop active load-pull, the output signal of the DUT is re-routed back after adjusting its phase and magnitude. The latter configuration has the downside of degrading the characterization system's inherent stability. Filters are generally introduced in the path of the signal in order to avoid oscillations. Also, the impedance provided by this configuration is limited over a narrow bandwidth. Thus, the open-loop active loading method is preferred, though it may suffer from lengthy iterative procedure to reach a specific impedance. On the other hand, in envelope load-pull, the output signal is first demodulated, then processed by an analog circuit before being again modulated and injected back into the DUT port.

Table 2.1 compare passive load-pull and active load-pull (open-loop and close-loop) under different criteria.

Table 2.1 - Comparison Between Passive Load-pull and Active Load-pull

	Passive Load Pull	Active Load Pull (Open Loop)	Active Load Pull (Closed Loop)
Inherent System Stability	Not prone to oscillations	Not prone to oscillations	Prone to oscillations
Coverage	Limited maximum reflection coefficient	All Smith Chart	All Smith Chart (linearity limit)
Impedance Synthesis	No iterations Independent on input drive levels	Iterations needed Dependent on input drive levels	No iterations Independent of input drive levels
Challenges	Difficult for high frequencies and high power	Difficult for high power	Difficult for high power
Bandwidth	Narrow-band	Broader bandwidth	Narrow-band
Harmonic Control	Easy	Difficult	Difficult
On-Wafer Measurement	Vibration Problem Additional leading loss	Easier	Easier

Cost	Cheapest	More Expensive	More Expensive

- *Hybrid load-pull:* Hybrid load-pull combines the passive load-pull and active load-pull in order to exploit the advantage of each load-pull type. These advantages are mainly the high power handling capabilities of passive tuners and Smith chart coverage of active load-pull. Hence, combining the two load-pull techniques should avoid the use of expensive drivers in the active load-pull and compensate for the losses in the passive load-pull.

One of the common techniques to achieve the combination of the active load-pull and passive load-pull is the use fundamental tuners instead of harmonic tuners to set the impedance at the fundamental frequency and to control actively the harmonics. Indeed, passive tuners at the fundamental frequency will reflect a part of the power to the DUT and hence avoid the use of expensive high-power drivers if the impedance is to be set actively. Moreover, the power at the harmonics are lower and hence active-load pull at the harmonics just need low-power broadband power amplifiers. Figure 2.11 presents an example of hybrid setup where the fundamental impedance is set actively and the impedance at the harmonics are set actively published in [6].

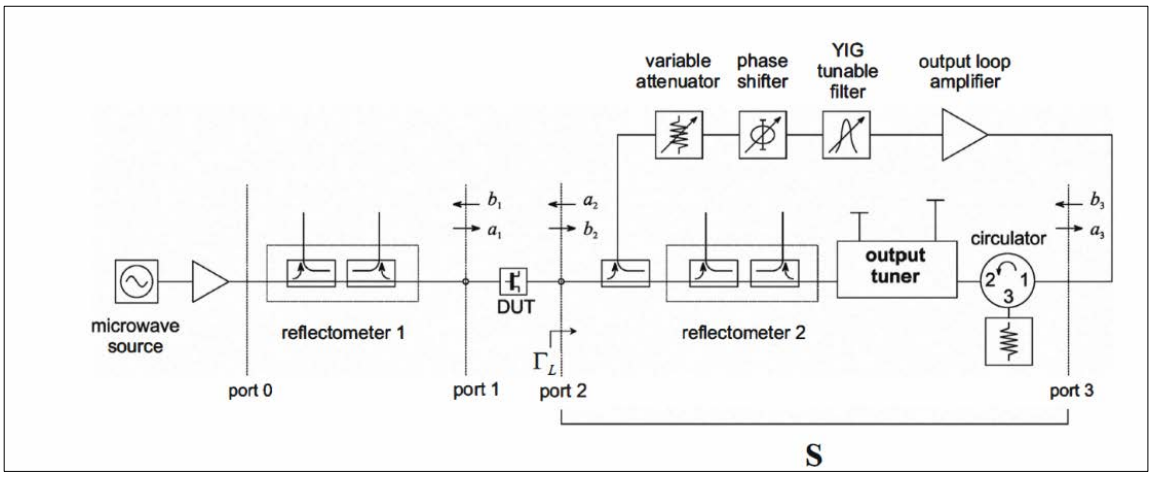


Figure 2.11 - Example of a Hybrid Load-pull [6]

2.2.2 State-of-the-art Nonlinear Characterization Setups

Passive load-pull the first load-pull systems to be available due to their simplicity of use. Pioneer work on active nonlinear characterization setups go back to the original contribution of Takayama who introduced the first open-loop active load-pull system in 1976 [7]. Afterwards, Bava and al. introduced the closed-loop active load-pull in [8]. Since then, several works have been conducted to increase the capacities of the load-pull setups. Each novel contribution focused in particular aspect to improve. The main aspects for which a nonlinear system should be evaluated are listed below:

- *Power Handling:* Power handling is the ability of the setup to characterize high power devices. This can be problematic especially in active load-pull as it requires expensive linear high-power drivers to reach the desired power levels and high load impedances.

- *Signal Bandwidth:* Signal bandwidth is important when the stimulus is a multitone signal or modulated signals.

- *Radio Frequency Range:* Frequency range determines the lowest and the highest frequencies that can be handled by the setup. Thus, it defines the number of harmonics that can be measured.

- *Stability:* Stability describes the possibility of the setup to induce oscillations.

- *Harmonic & Baseband Tuning Capability:* Nonlinear Setups are expected to control and to measure the signals at the DUT ports around the fundamental frequency but not all of them can control or measure the spectral components at the harmonics and low frequency, though arbitrary impedances at these frequencies can change radically the behaviour of the DUT.

- *Measurement Accuracy & Dynamic Range:* Errors in the characterization procedure degrades the performance in the modeling and can mislead the designer and delays the design process.

- *Reflection Coefficient Coverage:* As stated before, passive load-pull systems cannot cover the whole Smith chart and present a reflection coefficient equal to one. Active load-pull has no coverage limitation and can even present a reflection coefficient superior to one.

- *Device insight:* Device insight is mainly defined through the nature of stimulus used, the control over the device temperature and the spectrum components measured and time domain reconstruction ability.

The table below shows a compilation of the major developed setups available in the literature. Attempts have been made to develop characterization systems that satisfy the following criteria which are directly related to the targeted application by supporting specific types of stimulus, using particular waveform measurement instrument and appropriate waveform engineering approach.

Table 2.2 - Comparison of State-of-the-art Nonlinear Characterization Setups

Criteria		[6]	[9]	[10]	[11]	[12]
Waveform Measurement Instruments	Oscilloscope		x			
	LSNA				x	
	NVNA					x
	Other	x (1)		x (2)		
Waveform Engineering	Passive Load-pull (LP)					x
	Open loop LP				x	
	Close loop LP					
	Envelope LP		x	x		
	Hybrid LP	x				
Waveform stimuli	CW	x				x
	Pulsed CW				x	
	Multitone		x			
	Modulated Signal			x		
Frequency domain and Time Domain Capabilities	Harmonics	x		x	x	x
	Baseband			x	x	
	Time Domain				x	

(1) Vector network analyzer (VNA) ; (2) Mixer-based system

Figure 2.12 below summarizes of the different parts of nonlinear characterization systems in a tree-format where branch represent a possible setup configuration.

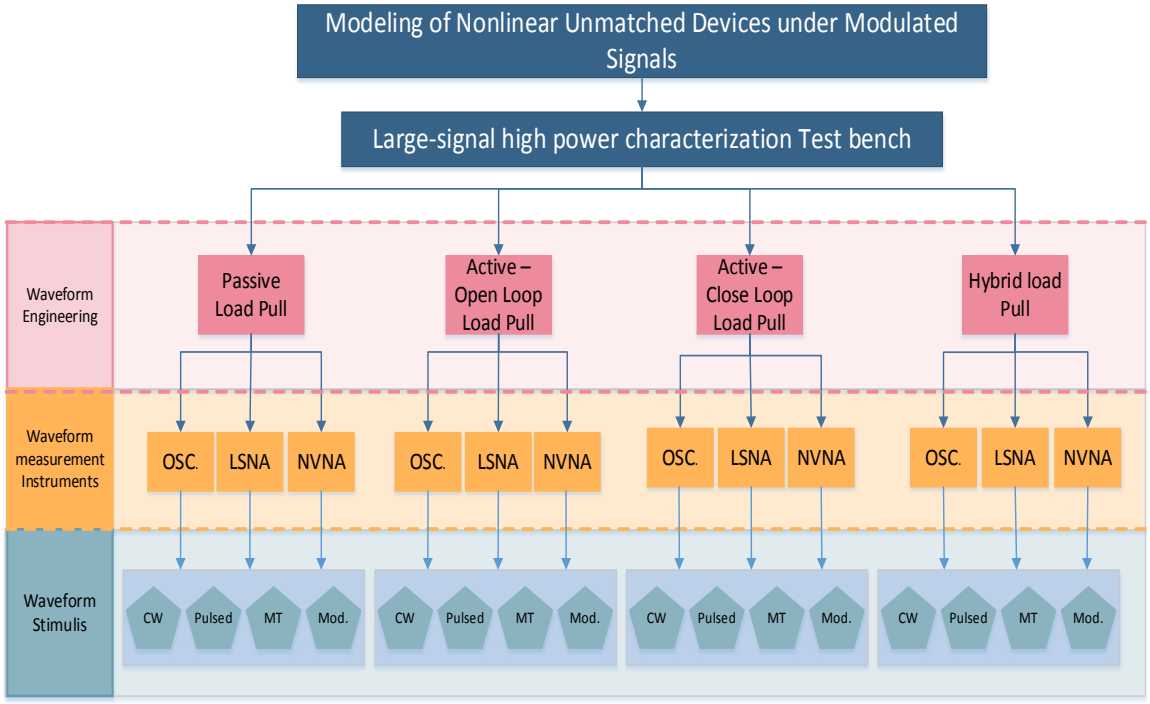


Figure 2.12 – Nonlinear Characterization Setup Structure Breakdown

2.3 Describing Functions Framework for Modeling Nonlinear Unmatched Devices

Within the describing functions framework, the quantities considered are the traveling voltage waves (or equivalently the power waves). Same as the case of S-parameters, they are the linear combinations of voltages and currents at the ports of the DUT (see Figure 2.1). A-waves (a_1, a_2) are the incident waves and B-waves (b_1, b_2) are the reflected waves as shown in equations (2.1) and (2.2).

$$a_i = \frac{V_i + Z_i\, I_i}{2\sqrt{Re(Z_i)}} \tag{2.1}$$

$$b_i = \frac{V_i - Z_i^{*}\, I_i}{2\sqrt{Re(Z_i)}} \tag{2.2}$$

Where *i* is the port index.

The variable Z_i is an arbitrary reference impedance and it has the value of 50 Ohms per default but it can vary depending on the DUT. For example, if the DUT is a transistor, Z_i can be as low as a few ohms to be close to its output impedance.

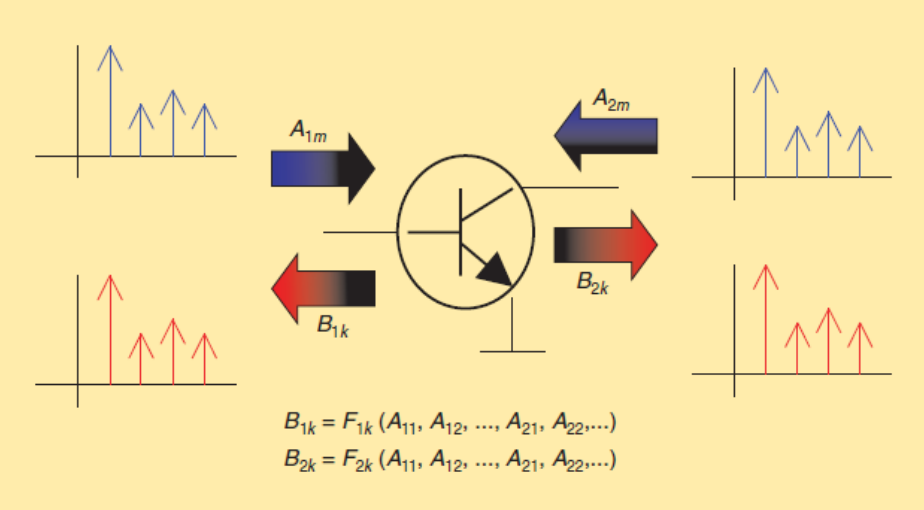


Figure 2.13 - Description Functions Representation [13]

In describing functions framework, the behaviour of DUT is modeled through harmonically related discrete multitone. It means that the spectral components of the input/output power waves are composed of the fundamental (f_0) and its harmonics ($2f_0, 3f_0$ ...). An immediate benefit of this formulation is its compatibility with modern commercial CAD where the signal at the ports of a nonlinear device is treated as a narrowband modulation around the carrier and its harmonics. The goal when using the describing functions framework is to find the frequency domain, nonlinear, multivariate, complex functions that map the output B-waves, at each harmonic and each port to the input waves A-waves fundamental and harmonics. Equation (2.3) describes this frequency domain mapping:

$$B_{pm} = F_{pm}(A_{11}, A_{12} \ldots, A_{21}, A_{22}, \ldots, A_{qn}, \ldots) \quad (2.3)$$

B_{pm} : Output power-wave at port p and harmonic m.

F_{pm} : Describing functions for B_{pm}.

A_{qn} : Input power-wave at port q and harmonic n.

Recently, two models have been proposed and showed interesting results using the describing functions formalisms: X-parameters and Cardiff model. They both involved a number of approximations and simplifications to the general equation (2.3) to measurement extraction procedure to build a complete model that could be incorporated in CAD environment. The details and limitations of each model are presented in the next two sub-sections.

2.3.1 X-parameters

X-parameters model is a particular case of equation (2.3). It is developed to be an extension of S-parameters to handle nonlinear systems. Indeed, X-parameters reduces to S-parameters when the DUT is operated in the small-signal regime. First, the formulation of X-parameters will be explained then the extraction procedure is described.

2.3.1.1 X-parameters Formulation

The full derivation of X-parameters is detailed in [13]. It is based on three assumptions:

The first assumption claims that the DUT is a time-invariant device which means that a delayed signal applied at the input of the DUT results on a delayed output. In the frequency domain, this propriety translates into a phase shift of the output proportional to the frequency. The

time-invariance propriety is used to assign a phase of zero to the fundamental tone A_{11} and normalize all the other frequencies to it in order to simplify the formulation. As illustrated in equation 2.4, the B-waves now only depend on the magnitude of A_{11}:

$$B_{pm} = F_{pm}\left(|A_{11}|, A_{12}P^2 \ \ldots, A_{21}P^2, A_{22}P^2, \ldots, A_{qn}P^n, \ldots\right)P^m \tag{2.4}$$

$Where\ P = e^{-j\angle A_{11}}$

The second assumption consists in the non-analytical nature of the nonlinear functions F_{pm}. This assumption results in introducing conjugate terms to the equation (2.3). These terms are not present in the case of S-parameters because linear functions are analytic. The results of introducing conjugate terms is shown in equation (2.5):

$$B_{pm} = F_{pm}\left(A_{11}, A_{12}\ , A^*_{12}\ldots, A_{21}, A^*_{21}, A_{22}, A^*_{22}, \ldots, A_{qn}, A^*_{qn}\ldots\right) \tag{2.5}$$

Finally, the X-parameters are based on the Harmonic Superposition Principle (HSP). The HSP sets the tones with the dominant power to define the Large Signal Operating Point (LSOP). In the case of a transistor or a power amplifier, The LSOP is generally defined by the magnitude of A_{11} which happens to be the main excitation. A_{21} is usually added to the LSOP when the output port of the DUT is unmatched resulting in a high reflection coefficient. The DUT behavior is linearized around each LOSP which means that the contribution of the other spectral components that do not belong to the LSOP (harmonics like A_{12} and A_{23}) are considered to be linear. For each LSOP variable, a different set of parameters is extracted. If it happens that the power of one of the harmonics is too high to be accounted for as a linear contributor, it is just added to the LSOP. The final formulation of the x-parameters is as follow:

$$\begin{aligned} B_{pm} = {} & X^F{}_{pm}\,(LSOP)\,P^m + \sum_{qn} X^S{}_{qn,pm}\,(LSOP)\,A_{qn}\,P^{m-n} \\ & + \sum_{qn} X^T{}_{qn,pm}\,(LSOP)\,A_{qn}{}^*\,P^{m+n} \end{aligned} \tag{2.6}$$

2.3.1.2 X-parameters Extraction

The extraction procedure of the X-parameters exploits the harmonic superposition principle which does not impose that the excitation signal has to be applied at the same time at all harmonics. Assume that the model to be extracted has an LSOP that just depend on $|A_{11}|$ and that only the harmonics A_{12} and A_{21} are considered. (see equation 2.7).

$$\begin{aligned} B_{pm} = {} & X^F{}_{pm}\,(|A_{11}|)\,P^m + X^S{}_{12,pm}\,(|A_{11}|)\,A_{12}\,P^{m-2} \\ & + X^S{}_{21,pm}\,(|A_{11}|)\,A_{21}\,P^{m-2} + X^T{}_{12,pm}\,(|A_{11}|)\,A_{12}{}^*\,P^{m+2} \\ & + X^T{}_{21,pm}\,(|A_{11}|)\,A_{21}{}^*\,P^{m+2} \end{aligned} \tag{2.7}$$

The first step of the extraction is to fix the LSOP. In the case of the model in (2.7), $|A_{11}|$ is held at constant power while extracting the other parameters: To extract the X^F term, only $|A_{11}|$ is applied and the harmonics are kept at zero. However, to extract the X^S and X^T terms, A_{21} is kept at zero and a perturbation, called tickler tone, is applied at A_{12} with different phase each time in order to yield enough equations to solve for the X^S and X^T terms related to A_{12} using equation (2.7). The same procedure is repeated where each time a tickler tone is applied at one harmonic and the remaining harmonics are set to zero (see Figure 2.13). When all the terms are extracted, the LSOP is changed to the next point in the sweep until the desired range is covered. Special

care has to be accorded to the power of the tickler tone because too low power may go below the noise floor and too high power may invalidate the harmonic superposition principle assumption since the harmonics then could exhibit nonlinear effect. (see Figure 2.14)

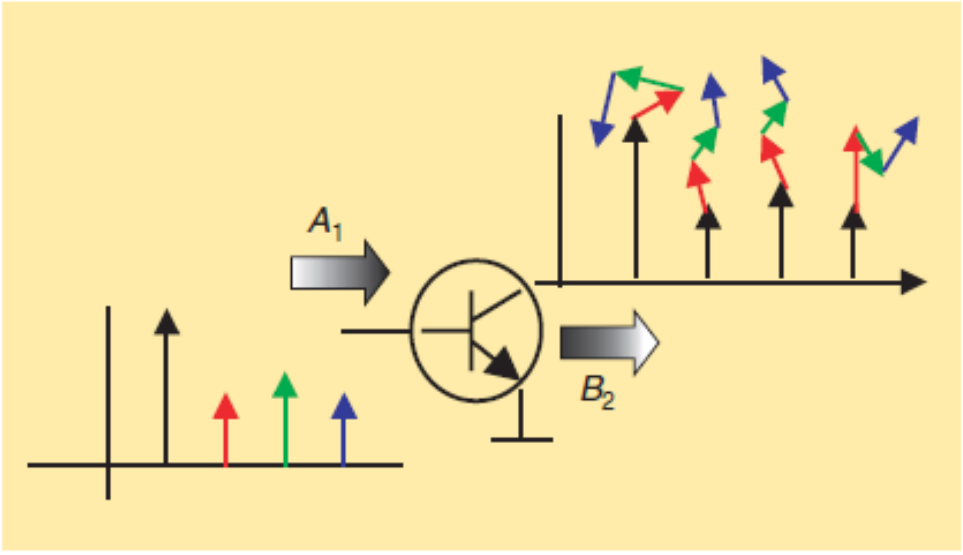


Figure 2.14 - Harmonic Superposition Principle [13]

The experimental setup to perform X-parameters extraction is based on the Agilent PNA-X [14]. The PNA-X measures incident and reflected waves in fundamental and harmonics frequencies both amplitude and phase. It is used in a combination with comb generator that present a phase reference at each frequency and ensure the phase coherency of the signal spectrum. Further details about the PNA-X based setup are available in chapter 4.

2.3.1.3 Generalization of X-parameters to Unmatched Devices

The X-parameters model under one tone stimulus as explained above is extracted for a limited impedance range typically around 50 Ohm. That means that once the load impedance is changed to a different value in the Smith Chart, it invalidates the X-parameters model if the behavior of the DUT changes nonlinearity with the load impedance. Hence, in [15], the X-parameters model has been augmented to account for variation of the load impedance. The new *load-dependent X-parameters* is able to present a model suitable for high mismatched environment such high-power and multi-stage power amplifiers and transistors that have their load impedance faraway for 50 Ohm.

Load-dependent X-parameters model is measured over a sweep of load impedance that can cover all of the Smith Chart or can be limited to a specific range that the user specifies. The load impedance is added as another parameter to the LSOP in addition to the DC bias and input drive power. For each load impedance, an X-parameters model is extracted and the results are combined into a single file and imported to a nonlinear simulator (Agilent ADS). In order to be able to vary the load impedance and extract x-parameters model, the initial extraction setup based on the Agilent PNA-X (with the NVNA software option) is augmented with a passive tuners that will allow to set the fundamental frequency at the desired reflection coefficient. (see Figure 2.15).

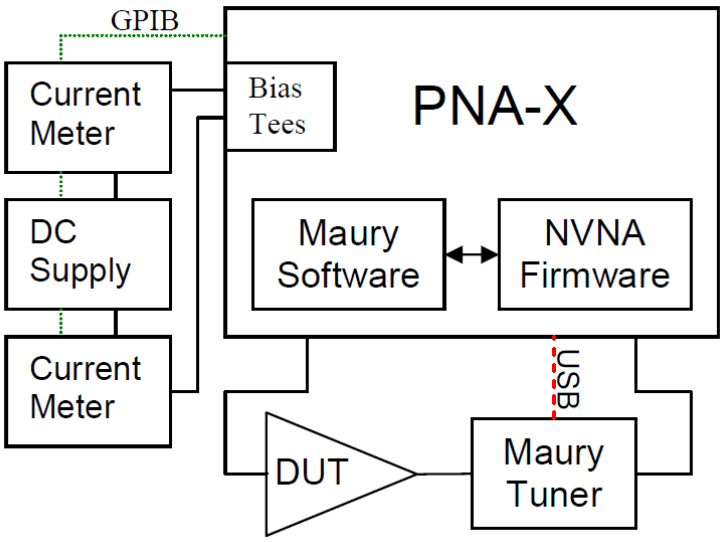


Figure 2.15 - Measurement Setup for Load-dependent X-parameters [15]

In [16], authors have performed further validation of the load-dependent x-parameters by comparing the results of the x-parameters model extracted with a load pull system to a commercially available model of the transistor from the manufacturer and they underlined the challenges associated with the extraction.

Over the last decade, X-parameters have known multiple enhancements that are not restricted to including the effect of load impedance. The major improvement is the attempt to include long-memory effects [17] since the initial formulation of X-parameters is inherently pseudo-static and can't model properly the dynamics of the DUT. The new formulation of the *dynamic X-parameters* introduces further complications to the model formulation and extraction procedure.

2.3.2 Cardiff Model

Using the describing functions framework, the Cardiff model is another attempt to formulate a practical and precise approximation function that can predict the output signal of DUT as a function of the input stimulus [18]. The signal considered are still the power waves in frequency domain. Restricting the input stimulus to fundamental frequency at port 1 ($A_{1,1}$) and the fundamental frequency at port 2 ($a_{2,1}$), the general equation (2.3) simplifies to the following:

$$B_{pm} = f_{pm}\left(A_{1,1}, A_{2,1}\right) \tag{2.8}$$

$where\ p = port\ index\ and\ m = harmonic\ index.$

2.3.2.1 Cardiff Model Formulation

While X-parameters model uses the cartesian representation of the input signals which treats the stimulus at the input ports of the DUT ($a_{1,1}$ and $a_{2,1}$) as complex number and their conjugates, the Cardiff model uses the polar representation where the stimulus are represented in terms of their magnitude and phase. Moreover, the absolute phase of the stimulus at the fundamental frequency at port 1 (phase of $a_{1,1}$) is used as reference and only the relative phase is considered. The resulting equation is showed below:

$$B_{pm} = {P_1}^m\left(|A_{1,1}|, |A_{2,1}|, \frac{Q_1}{P_1}\right) \tag{2.9}$$

where $P_1 = e^{-\angle A_{1,1}}$ and $Q_1 = e^{-\angle A_{2,1}}$.

Equation (2.9) demonstrates the independence between the phase and the magnitude of the input stimulus. Hence, the effect of each parameter can be modeled and extracted separately from the other. This independence is the central aspect that the Cardiff model is based on. In fact, the output signals are expected to be subject to periodic changes with respect to the phase if the magnitude of the input signals is to be kept constant. Consequently, the authors in [19] used the Fourier series to model the dependency on phase resulting in equation (2.10).

$$B_{pm} = {P_1}^m \sum_{n=-(N-1)/2}^{(N+1)/2} \left\{ R_{p,m,n}\left(|A_{1,1}|, |A_{2,1}| \right) \left(\frac{Q_1}{P_1} \right)^n \right\} \tag{2.10}$$

where N = order of the model.

The separation of phase and magnitude simplifies the measurement process to extract the model coefficients. $R_{p,m,n}$ coefficients are functions of the magnitude of the input and output stimulus in both ports and can be represented by polynomials with an order no larger than the order identified. To capture the complete phase response at a given power level, the power level of injected tones at the fundamental at both ports are kept at constant magnitude while their relative phase is swept through 360 degrees.

The model presented in (2.9) accounts only for the excitation at the fundamental frequency at both port. To be able to model properly a DUT like a power amplifier or a transistor, the harmonics have to be accounted for correctly in order to choose the right terminations during the design process. Harmonics terminations of power amplifier significantly affect its performances. Moreover, the harmonic terminations of certain class of operation like E, J, J^{-1} ... is well-defined. Unlike X-parameters, Cardiff model does not assume that the contribution of the harmonic is linear around the LSOP. Instead, the same approach for the fundamental frequency is expanding to harmonics by including the power and relative phase of the harmonics into the model. A model that include the effect of $a_{2,2}$ is presented in equation (2.11).

$$B_{pm} = {P_1}^m \sum_n \sum_r \left\{ G_{p,m,n,r}\left(|A_{1,1}|, |A_{2,1}|, |A_{2,2}| \right) \left(\frac{Q_1}{P_1} \right)^n \left(\frac{Q_2}{{P_1}^2} \right)^r \right\} \tag{2.11}$$

If other harmonics are to be added to the model, the same approach is used again. The drawback of this approach is that the number of coefficients and measurements will increase exponentially each time a new harmonic is added to the formulation. It is to be noticed that the fact the $a_{2,1}$ as well as $a_{2,2}$ effect is modeled in terms of magnitude and phase, the Cardiff model is inherently able to capture the behaviour of devices in an unmatched environment.

2.3.2.2 Cardiff Model Extraction

The measurement procedure for extracting the model in equation (2.10) involves a set of nested sweeps of magnitude and phase. First, a sweep of magnitudes of the fundamental at both port 1 and 2 as well as the second harmonic stimuli at port 2 are performed. For each point of this sweep, i.e. the triplet ($|a_{1,1}|, |a_{2,1}|, |a_{2,2}|$), the phase of fundamental input stimuli at port 2 (phase of $a_{2,1}$) is swept around 360 degrees, with discrete steps. Finally, for each phase of $a_{2,1}$, the phase of the second harmonic at port 2 (phase of $a_{2,2}$) is stepped around the 360 degrees. The waveform measurement setup used to extract the Cardiff Model was developed at Cardiff University [20].

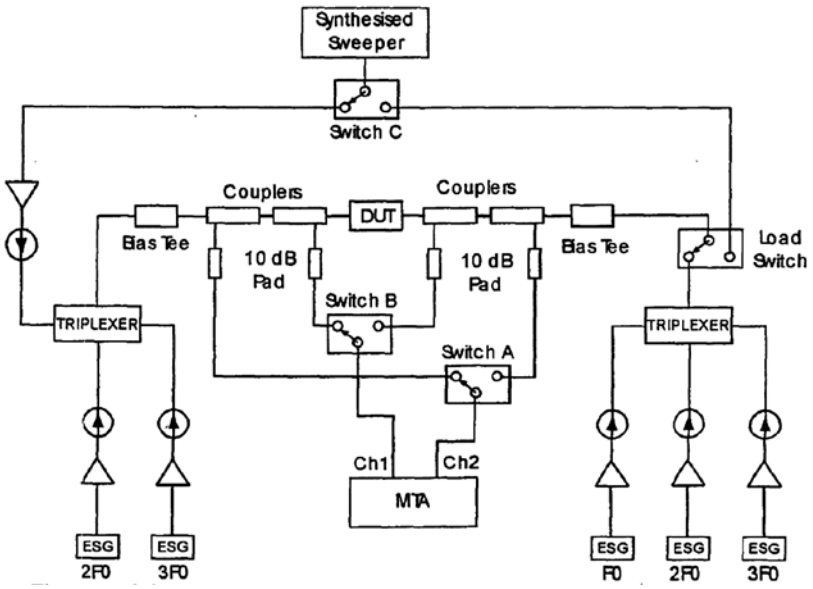


Figure 2.16 - Two-port Time Domain Measurement System [21]

The measurement data is renormalized to the measured system impedances and the fundamental phase at port 1 is removed since all phase are referenced to the phase to the fundamental frequency at port 1. The measurement system to extract the model is achieved through combining phase coherent signal generators (ESG). Each signal generator produces a CW with the desired phase and magnitude at the fundamental or harmonics. For each additional harmonic to the setup, another ESG is needed. The receivers consists of Microwave Transistor Analyser (HP-MTA). However, the setup for extracting Cardiff model has been upgraded over the years to include the Tektronix AWG instead of the combined ESGs and the Tektronix oscilloscopes as receiver to reach higher frequency and offers more flexibility.

Once the data is collected, the identification of the coefficients could be performed through two possible approaches. The first approach consists in treating equation (2.11) as a Two-Dimensional Fourier transform. Using Fourier transform allows all the coefficients to be extracted at once without the risk of over-fitting the data, but suffers from errors when the sample data are noisy. In the other hand, an iterative algorithm could be used like the least mean squares (LMS) method. LMS could result in over-fitting but does not suffer from leakage problems.

2.4 Volterra Series Framework for Nonlinear Matched Devices

2.4.1 Volterra Series Formulation

Linear devices can be completely characterized by their impulse response. In the case of nonlinear devices, memoryless polynomials were used as well as look-up tables to capture the nonlinearity of matched devices. However, the increase in the bandwidth of the modern modulated signals (> 100 MHz) has accentuated the contribution of the memory effects in the signal distortions where the dynamics behavior of the device have to be considered in addition to its pseudo-static behavior.

Different modeling frameworks have been subject to intensive research mainly involving Volterra Series and ANN due to their abilities to model nonlinearities with fading memory effect. Volterra Series offer a systematic representation of time-invariant systems and give relatively more insight on the nonlinear characteristics compared to ANN which is inherently a blind process [22]. The Volterra series can be considered as Taylor series expansion that account for fading memory effect. Equation (2.12) shows Volterra Series written in Time domain:

$$y(t) = \sum_{n=1}^{\infty} \int h_n(\tau_1, \dots, \tau_n) \prod_{p=1}^{n} x(t - \tau_p) d\tau_p \tag{2.12}$$

Where: $y(t)$ is the output signal and $x(t)$ is the input signal.

In addition to the system's impulse response, the Volterra series includes a multidimensional convolution integrals with increasing order. The terms $h_n(\tau_1, \dots, \tau_n)$ are called the n-th order *Volterra kernels.* The series convergence may require a very large number of kernels to be included. The frequency domain Volterra series is illustrated in equations (2.13) and (2.14). The spectrum of the output is the combined mixed-products of the input spectral components resulting in new frequencies. The coefficients $H_n(f_1, \dots, f_n)$ are the multi-dimensional Fourier transform of the time domain Volterra kernels and are called the n-th order transfer functions.

$$y\,(f) = \sum_{n=1}^{\infty} y_n(f) \tag{2.13}$$

$$where$$

$$y\,(f) = \iiint H_n(f_1, \dots, f_n)\ \delta(f - f_1 - \cdots f_n) \prod_{i=1}^{n} X(f_i)\, df_i \tag{2.14}$$

When applied in the context of system level simulations and modeling of the front-end transceivers components, often the truncated version of Volterra series is used in the envelope domain. (See equations 2.15 and 2.16).

$$\hat{y}\,(k) = \sum_{n=1}^{N} y_n(k) \tag{2.15}$$

$$where$$

$$y_n(k) = \sum_{i_1=1}^{M-1} \dots \sum_{i_n=1}^{M-1} h_n(i_1, \dots, i_n) \prod_{p=1}^{n} \hat{x}(k - i_p) \tag{2.16}$$

In equations 2.14 and 2.15, $\hat{y}\,(k)$ and $\hat{x}\,(k)$ are the baseband signal (I/Q signal) that contains the information that is transmitted. The modulated signal transmitted is the up-converted version of the baseband signal around the carrier frequency. (See equation 2.17)

$$x(t) = \hat{x}\,(t) \cos(2\,\pi\, f_c\, t) \tag{2.17}$$

Since RF power amplifiers are the major component that dictates the linearity and efficiency of transceivers, Volterra series has been used extensively to model the RF power amplifiers behavior as well as finding a predistorion algorithm to compensate for its inherent nonlinearity.

2.4.2 Popular Volterra Series Derivations

The complexity of the Volterra series increases dramatically especially when modeling hard nonlinearity. For this reason, the use of Volterra series has long been restricted to weak nonlinearity where the order of linearity does not exceed 3rd or 5th order. To overcome the problem of complexity, a lot of work has been done to derive less complex model from the

original Volterra series with acceptable accuracy. The general idea is to reduce the number of cross-terms in intelligent and systematic way where cross-terms are the terms that involves multiple instances of the input with different delays.

2.4.2.1 Memory Polynomial

One of the most successful derivations of Volterra series is the Memory Polynomial first introduced in [23]. The Memory Polynomial does not include the cross terms from the Volterra series formulation and keeps only the power of the input samples. Equation (2.13) simplifies to the following:

$$\hat{y}_{MP}(k) = \sum_{n=0}^{N} \dots \sum_{m=1}^{M} h_{m,n}\ \hat{x}(k-m)|\hat{x}(k-m)|^{m-1} \tag{2.18}$$

Where $\hat{y}_{MP}(k)$ is the envelope of the output signal and $\hat{x}(k)$ is the envelope of the output signal, while N is the order of nonlinearity and M is the memory length.

The formulation of the Memory polynomial is linear with respect to its kernels. Thus, the kernels can be extracted using the least square estimator (LSE) algorithm. Because of its reduced complexity as well as the modeling capability, the Memory polynomial is considered as a golden reference when comparing models based on Volterra Series.

2.4.2.2 Dynamic Deviation Reduction-Based Volterra series

Dynamic deviation reduction-based Volterra series (DDR Volterra series) is another derivation of Volterra Series. It is a trade-off between the original Volterra series and the Memory Polynomial in terms of accuracy and complexity. DDR Volterra does not eliminate all the cross terms as the Memory polynomial does, but it keeps a limited number of cross terms achieving more accuracy while still being less complex than the Volterra series.

The formulation is derived in [24] and the new model formulation is expressed below:

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} h_{p,0}(0,\dots,0)x^{p}(n) \\ &+ \sum_{p=1}^{P}\left\{\sum_{r=1}^{p}\left[x^{p-r}\sum_{i_1=1}^{M}\dots\sum_{i_r=i_r-1}^{M} h_{p,r}(0,\dots,0,i_1,\dots,i_r)\prod_{j=1}^{r}x(n-i_j)\right]\right\} \end{aligned} \tag{2.19}$$

The DDR Volterra separates the static response from the dynamic response in the model formulation. Moreover, the advantage of the DDR Volterra model over other model-pruning techniques is the fact that the coefficient reduction keeps the linearity with respect to the model coefficients allowing for LSE algorithm or any other well-established linear estimation algorithm to still be used. The number of coefficients in the DDR Volterra increases almost linearly with respect to nonlinearity order and memory depth, while this increase used to be exponential in the traditional Volterra series.

2.4.3 Nonlinear Device Behavioral Models Comparison

A full comparison is detailed between the X-parameters model and the Cardiff model in terms of performances and extraction procedures in [25]. Thus, this section presents a comparison between the models developed using the describing functions framework and Volterra Series derived models.

The major differences observed between these two types of models are summarized below:

- *Matching Considerations:* SISO Volterra Series presented above are applied for matched conditions. They cannot be applied for unmatched conditions unlike X-parameters and the Cardiff model do.
- *Nature of the stimulus:* In the Volterra Series framework, the used signals are modulated signals as they model system level components in real application while the models using the describing functions framework use CW signals for extraction. Thus, the models extracted under CW could not accurately capture the behavior of the device under modulated signals.
- *Signal Representation Domain:* Most of the work involving Volterra Series model the devices in the envelope domain or Time Domain. However, the describing functions framework represent the models in the Frequency domain. One advantage of the frequency domain representation is that it is easier to implement in commercial simulator like Harmonic Balance.
- *Memory Effects:* Volterra Series inherently account for memory effects. However, accounting for the memory effect in the Cardiff model and X-parameters is not straightforward and complicates significantly the formulation [15].
- *Number of Measurements:* Since its extraction is based on modulated signals, extracting Volterra series models involves considerably less measurements. In the case of X-parameters and the Cardiff model, when the number of variables defining the measurements conditions increases, the number of measurements could become impractical.
- *Complexity:* The main drawback of Volterra series is its complexity which is the origin of many model-pruning attempts. X-parameters and Cardiff models are less complex. X-parameters could be even thought to be a first order approximation of Volterra series. However, this simplification reduces the generality of the describing functions models making them local measurement-based models.

Chapter 3

DIDO Volterra Model Formulation and Simulation

Based on the previous comparison and discussions presented in chapter 2, adopting Volterra Series to unmatched environment to model two-port network under modulated signals will enable predicting the quasi-memoryless as well as the memory effects exhibited by the DUT. However, the model should have manageable complexity for practical implementation in CAD environments. This chapter discusses the Volterra model derivation and simulation, while chapter 4 is dedicated to the characterization system needed to extract the model. First, the steps followed to derive DIDO Volterra are detailed in section 3.1. Then, the simulation based validation and the corresponding results are presented in section 3.2.

3.1 DIDO Volterra Series Formulation

3.1.1 Initial DIDO Volterra Formulation

The SISO Volterra formulation in continuous time is recalled in equation 3.1.

$$y(t) = \sum_{n=1}^{\infty} \int h_n(\tau_1, \dots, \tau_n) \prod_{p=1}^{n} x(t-\tau_p) d\tau_p \tag{3.1}$$

In the case of matched nonlinear device, $y(t)$ represents the output signal and $x(t)$ represents the input signal. Hence, to generalize this formulation to the case of two-port unmatched network, the outputs and the inputs are composed of a vector of two signals (see equation 3.2).

$$\begin{gathered} Y = fct(X) \\ Y = [y_1, y_2]\,;\; X = [x_1, x_2] \end{gathered} \tag{3.2}$$

The two outputs, y_1 and y_2, are independent from each other but each output is the result of nonlinear combination and interaction of the two inputs, x_1 and x_2. For a two-port network, the input and output signals can be voltages and currents at the ports of the device. As transistors are voltage-controlled current-source devices, the currents should be modeled as a function of the voltage for seamless integration in CAD tools. However, the input and output signals can also refer to the power waves where the B-waves represent the output Y and the A-waves represent the input X. Since the characterization system measures the power waves instead of currents and voltages, the power waves representation was preferred in this work. Note that both representations are equivalent as a linear relation exists between them given the characterization impedance of the system (see equation 2.1 and 2.2). The generalization of the continuous-time SISO Volterra model to DIDO Volterra model is illustrated in equation 3.3.

$$Y_i(t) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} y_{i,mn}(t) \tag{3.3}$$

where

$$y_{i,mn}(t) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h_{i,mn}(\tau_{11}, \dots, \tau_{1m}; \tau_{21}, \dots, \tau_{2n}) \prod_{r=1}^{m} x_1(t - \tau_{1r}) d\tau_{1r} \prod_{s=1}^{n} x_2(t - \tau_{2s}) d\tau_{2s}$$

$Y_i(t)$ represents the output of the i^{th} port of the network. Similarly, a multiple-input multiple-output (MIMO) Volterra model can be derived for N port systems at the cost of higher complexity and more difficult extraction procedure.

Handling the instantaneous waveforms as shown in equation 3.3 would require an impractical sampling rate. Common CAD simulator supporting modulated signals handle instead the envelope domain representation of the signal as the information transmitted through the telecommunication path resides in the baseband signal and is independent of the carrier frequency. The DIDO Volterra model of equation 3.3 is transformed to the envelope domain as per equation 3.4

$$Y_i(k) = \sum_{n_1=1}^{N_1} \dots \sum_{n_2=1}^{N_2} y_{i,n_1 n_2}(k)$$

$$\begin{aligned} & y_{i,n_1 n_2}(k) \\ & = \sum_{i_1=0}^{M_1-1} \dots \sum_{i_{n1}=0}^{M_1-1} \sum_{j_1=0}^{M_2-1} \dots \sum_{j_{n2}=0}^{M_2-1} h_{i,n_1 n_2}(i_1, \dots, i_{n_1}; j_1, \dots, j_{n_2}) \prod_{p=0}^{M_1-1} x_1(k \\ & - i_p) \prod_{s=0}^{M_2-1} x_2(k - j_s) \end{aligned} \tag{3.4}$$

where N_1 and N_2 represent the two nonlinearity order and M_1 and M_2 represent the two memory depths associated with each input. Unfortunately, no systematic method exists for setting the nonlinearity order and memory depth of the model and it depends mainly on the device under test. For example, if a transistor with low reflection coefficient at the load is modeled, the model parameters of the input at port 1 of the device are expected to be larger than their counterparts of the waveform reflected at port 2. Note also that each output port may involve very different linearity order and memory depth from the two inputs.

Unlike the case of the SISO Volterra series formulation where all the kernels satisfy the propriety of symmetry, the symmetry of the DIDO Volterra kernels is port-dependent. The kernels symmetry means that all the kernels that only differ by a permutation of their arguments are equal and can be effectively represented by just one kernel representing the average of all other possible permutations. Consider the example of the SISO Volterra model where, for a given indexes i_1, i_2 and i_3, the kernels $h_3(i_1, i_2, i_3)$, $h_3(i_2, i_1, i_3)$, $h_3(i_1, i_3, i_2)$ and $h_3(i_3, i_1, i_2)$ are equal and can be represented by one unique kernel. For the DIDO Volterra case, $h_{i,22}(i_1, i_2; j_1, j_2)$ and $h_{i,22}(i_2, i_1; j_1, j_2)$ are equivalent (since arguments related to the same port are exchanged), meanwhile, $h_{i,22}(i_1, i_2; j_1, j_2)$ and $h_{i,22}(j_1, i_2; i_1, j_2)$ are not. As symmetry across port cannot be guaranteed for DIDO models, the complexity burden of the resulting model is more pronounced.

Despite the significant increase of the number of kernels with the new formulation, the proposed model maintains the linear property of the Volterra series with respect to kernels. This enables using the linear and optimal LSE.

3.1.2 DIDO Volterra Complexity Reduction

The complexity of Volterra models is measured using the number of kernels used in the formulation. Thus, if the nonlinear device exhibits strong nonlinearity or/and significant memory

effects, the number of kernels needed to ensure the convergence of the series will be unpractical. As a result, the use of Volterra series in SISO scenarios was restricted to mild nonlinearity and an extensive research on pruning Volterra Series was published aiming at reducing its complexity at the cost of its accuracy. DIDO Volterra has first to solve the complexity issue hindering the development of Volterra formulation for DIDO systems with acceptable performance. In the following, the increase of complexity associated with the generalization of SISO formulation to DIDO formulation is presented. Later, the methodology to reduce the complexity is explained and an example is provided.

Equation 3.5 presents the number of kernels for the case of SISO Volterra series for a given order of nonlinearity N and a memory effect M. This formula is applicable to the original Volterra Series without any pruning simplifications. Note that the number of kernels grows exponentially when the memory order or the nonlinearity order increases.

$$\sum_{n=1}^{N} \binom{n+M-1}{n} = \sum_{n=1}^{N} \frac{(n+M-1)!}{n!\,(M-1)!} \tag{3.5}$$

Equation 3.6 presents the number of kernels for the DIDO Volterra model as a function of the memory depth (M_1, M_2) of each input and a total nonlinearity order N. In fact, the nonlinearity order of each input depth (N_1, N_2) is not treated independently in this formula. Instead, the total nonlinearity order N is used which means that only the combinations of $N_1$and N_2 that sum to N are considered.

$$\sum_{n_1+n_2=1}^{N} \binom{n_1+M_1-1}{n_1}\binom{n_2+M_2-1}{n_2} = \sum_{n_1+n_2=1}^{N} \frac{(n_1+M_1-1)!}{n_1!\,(M_1-1)!}\ \frac{(n_2+M_2-1)!}{n_2!\,(M_2-1)!} \tag{3.6}$$

To further illustrate the problem of increased complexity for DIDO Volterra model when compared to SISO Volterra model, Table 3.1 shows the number of kernels when the order of nonlinearity and the memory effect is swept from 1 to 5. In the case of DIDO Volterra model, N is the total nonlinearity order ($N_1 + N_2 = N$) while both the memory depth of each input is equal to N ($M_1 = M_2 = M$).

Table 3.1 - Number of Kernels in SISO Volterra Model and DIDO Volterra Model

	SISO	**DIDO**
N=1; M=3	3	6
N=3; M=3	19	83
N=5; M=3	55	461
N=5; M=5	**251**	**3002**

As a result, particular care has to be given to reduce the complexity of the model without affecting its precision. The procedure to reduce the complexity of the DIDO Volterra model is based on two complementary methods. First, one should limit the combinations in terms of nonlinearity order. Secondly, the process developed by B. Fehri and S. Boumaiza in [26] to

reduce the complexity of the Volterra series in order to elaborate a digital predistorter for RF power amplifiers linearization is adopted to the purpose of unmatched nonlinear device modeling.

3.1.2.1 Nonlinearity Order Combinations

The DIDO Volterra model is a dual-input dual-output scheme modeling a two-port network and it is defined by four parameters: the nonlinearity order of each input (N_1, N_2) and the memory depth of each input (M_1, M_2). (see equation 3.7)

$$[\mathrm{y_1(t), y_2(t)}] = \mathrm{DIDO[x_1(t),\ x_2(t); M_1, N_1, M_2, N_2}] \tag{3.7}$$

In its most general form, any combination (n_1, n_2) between the nonlinearity order of the first input and the second input should be included. The combinations that have the nonlinearity of one of the inputs equals to zero, i.e. ($n_1, 0$) or ($0, n_2$), describe the impact of each input on the output independently from the other input; while, the remaining combinations describe the coupled effects and interaction of both inputs. Some combinations might have more impact in the model performance than other, but there is no systematic method to determine which combination to include and which combination to discard. This depends on the dynamics of the device under test. Note that the same analysis applies to the memory depth parameters (M_1, M_2).

Due to the significant complexity of the model, accounting for all possible combinations at the output is not practical. Thus, limiting these combinations is important to achieve reasonable complexity. To our knowledge, there is no generic theoretical technique to proceed with this reduction and any effort in this direction is empirical and iterative based on the prior knowledge about the DUT and findings through experiment results. For example, if the device under test is a transistor or an RF power amplifier, it is expected that the nonlinearity order for the input at port 1 is higher than the nonlinearity order for the input at port 2 as the dynamics of the device depends mainly on the primary excitation at port 1 and depends with lower degree on the reflected signal at port 2. If the reflection coefficient is low, the impact of the reflected power-wave at port 2, which constitutes the second input, could be thought of as a perturbation to the main excitation. Thus, as a guideline when instigating the model parameters N_1 should be set to be lower or equal to N_1.

In this work, an alternative method has been used to reduce the number of combinations by defining a *total nonlinearity order* and use that parameter instead of using the individual nonlinearity order of each input. As introduced previously, the total nonlinearity order N is the sum of N_1 and N_2 . For example, if N is equal to 3, the only combinations of (n_1, n_2) allowed to be included in the model are (0, 3), (1, 2), (2, 1) and (3, 0). The harder the nonlinearity of the device under test is, the higher N will be. Manipulating N should be easier that manipulating two nonlinearity parameters especially that SISO Volterra model can be used to give an approximation of the total nonlinearity order N. In fact, when moving from a matched environment to an unmatched environment, introducing a second input should capture the effect of varying the matching conditions and consequently avoid that the model require very high nonlinearity to account for its new dynamics. The total nonlinearity order of DIDO Volterra model should be in the same order of the SISO Volterra model.

3.1.2.2 Kernels Number Reduction

In order to reduce the number of kernels in the DIDO Volterra model, the technique used in [26] is adopted. The main important benefit of using this method is the *kernel sharing propriety* where the model effectively combines a number of basis functions to share a unique kernel. The steps for applying this derivation are highlighted in Figure 3.1.

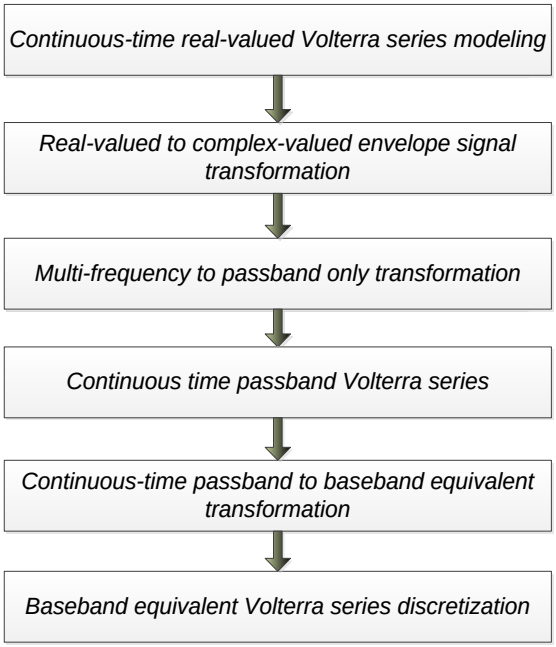


Figure 3.1 - Baseband Equivalent Modeling Approach [26]

Starting from the continuous-time formulation of the DIDO Volterra model, the discrete-time DIDO Volterra model, in the envelope domain, is extracted through six steps:

1. *Continuous-time real-valued Volterra series modeling:*

The first step in the derivation is to use the continuous-time DIDO Volterra formulation in equation (3.3) to describe the mapping between the inputs $[\ x_1(t), x_2(t)\ ]$ and outputs $[\ y_1(t), y_2(t)\ ]$ of the DUT. The signals used in this formula are the RF signals around the carrier frequency and not the envelope signals.

2. *Real-valued to complex-valued envelope signal formulation:*

The input signals $[\ x_1(t), x_2(t)\ ]$ are assumed to be single harmonic. Therefore, $[\ x_1(t), x_2(t)\ ]$ are treated as band-limited modulated signals where the envelope signals that contain the information is up-converted to the carrier frequency, can be reformulated as in equation (3.8) and (3.9).

$$\mathrm{x_1(t)} = \mathrm{Re}\{\hat{\mathrm{x}}_1(\mathrm{t})\mathrm{e}^{\mathrm{jw_c t}}\} = \frac{1}{2}\left(\hat{\mathrm{x}}_1{}^*(\mathrm{t})\mathrm{e}^{-\mathrm{jw_c t}} + \hat{\mathrm{x}}_1(\mathrm{t})\mathrm{e}^{\mathrm{jw_c t}}\right) \tag{3.8}$$

$$\mathrm{x_2(t)} = \mathrm{Re}\{\hat{\mathrm{x}}_2(\mathrm{t})\mathrm{e}^{\mathrm{jw_c t}}\} = \frac{1}{2}\left(\hat{\mathrm{x}}_2{}^*(\mathrm{t})\mathrm{e}^{-\mathrm{jw_c t}} + \hat{\mathrm{x}}_2(\mathrm{t})\mathrm{e}^{\mathrm{jw_c t}}\right) \tag{3.9}$$

Hence, replacing $x_1(t)$ and $x_2(t)$ in equation (3.3) results in an output spectrum that contains band-limited signal around DC, fundamental frequency and harmonics as the device nonlinearity creates new frequency components (see equation 3.10).

$$\mathrm{y_i(t)} = \sum_{\mathrm{p=-N}}^{\mathrm{N}} \frac{1}{2}\left(\hat{\mathrm{y}}_\mathrm{p}{}^*(\mathrm{t})\mathrm{e}^{-\mathrm{jpw_c t}} + \hat{\mathrm{y}}_\mathrm{p}(\mathrm{t})\mathrm{e}^{\mathrm{jpw_c t}}\right) \tag{3.10}$$

Where y_i is the output of the i^{th} port and N represent the number of harmonics.

The new spectrum content is shown in *Figure 3.2*. As the two inputs are around the same carrier frequency, the output spectrum does not have spectral components outside the baseband and harmonic frequencies. The band-limited signal around each harmonic frequency is treated as an independent signal and is written as a function of the two input envelopes.

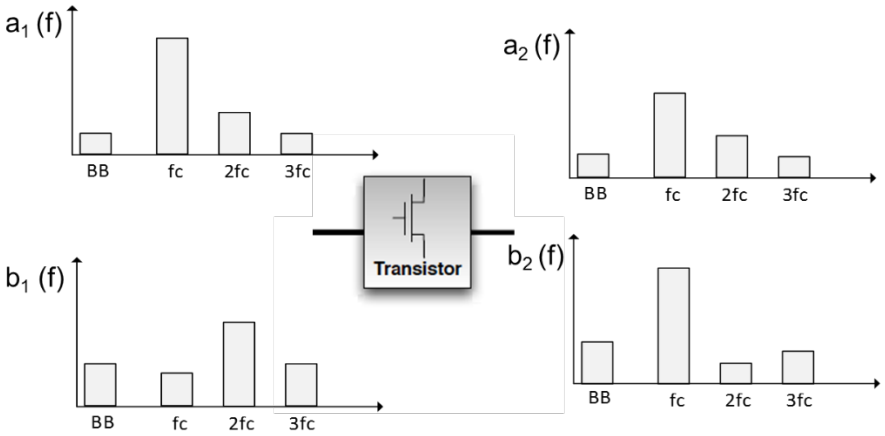


Figure 3.2 - Nonlinear Device Outputs Frequency Spectrum

3. *Multi-frequency to passband only transformation:*

Out of all the spectrum components that constitute the output signals, only the terms that describe the envelope of the output around the fundamental frequency are selected. It comes down to just considering the terms where p=1 in equation (3.10) which results in the simplified equation (3.11).

$$y_{i,pb}(t) = \frac{1}{2}\left(\hat{y}_1{}^*(t)e^{-jw_c t} + \hat{y}_1(t)e^{jw_c t}\right) \tag{3.11}$$

4. *Continuous-time passband Volterra series:*

Considering only the output response around the fundamental frequency, the DIDO model is rewritten in equation (3.12) as the simulation of nonlinear order response. Note that only the odd nonlinearity order are included since the even nonlinearity order does not generate new frequencies around the fundamental frequency.

$$\hat{y}_1(t)e^{jw_c t} = \sum_{k=0}^{\infty}\left(y_{1,2k+1}(t)\right) = \left(\sum_{k=0}^{\infty} y_{1,2k+1}(t)\right)e^{jw_c t} \tag{3.12}$$

For illustration purpose, the 3rd order nonlinearity is shown in equation (3.13)

$$
\begin{aligned}
y_{1,3}(t) = &\iiint_{-\infty}^{+\infty} h_{3,0}(\tau_1,\tau_2,\tau_3)\ BF_{3,0}(x_1(t),x_2(t),\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3 \\
&+ \iiint_{-\infty}^{+\infty} h_{0,3}(\tau_1,\tau_2,\tau_3)\ BF_{0,3}(x_1,x_2,\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3 \\
&+ \iiint_{-\infty}^{+\infty} h_{1,2}(\tau_1;\tau_2,\tau_3)\ BF_{1,2}(x_1,x_2,\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3 \\
&+ \iiint_{-\infty}^{+\infty} h_{2,1}(\tau_1,\tau_2;\tau_3)\ BF_{2,1}(x_1,x_2,\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3
\end{aligned}
\tag{3.13}
$$

$BF_{m,n}(x_1, x_2, \tau_1, \dots, \tau_{m+n})$ are called *basis functions*. They are the combinations of *m* instances of the first input and *n* instances of the second input associated with each kernel. For example, equation (3.14) shows the expression for $BF_{1,2}(x_1, x_2, \tau_1, \tau_2, \tau_3)$.

$$
\begin{aligned}
&BF_{1,2}(x_1,x_2,\tau_1,\tau_2,\tau_3) \\
&\quad = \left(\hat{x}_1(t-\tau_1)e^{jw_c(t-\tau_1)}\right)^*\left(\hat{x}_2(t-\tau_2)e^{jw_c(t-\tau_2)}\right)\left(\hat{x}_2(t\right. \\
&\quad \left.-\tau_3)e^{jw_c(t-\tau_3)}\right) \\
&\quad + \left(\hat{x}_1(t-\tau_1)e^{jw_c(t-\tau_1)}\right)\left(\hat{x}_2(t-\tau_2)e^{jw_c(t-\tau_2)}\right)^*\left(\hat{x}_2(t\right. \\
&\quad \left.-\tau_3)e^{jw_c(t-\tau_3)}\right) \\
&\quad + \left(\hat{x}_1(t-\tau_1)e^{jw_c(t-\tau_1)}\right)\left(\hat{x}_2(t\right. \\
&\quad \left.-\tau_2)e^{jw_c(t-\tau_2)}\right)\left(\hat{x}_2(t-\tau_3)e^{jw_c(t-\tau_3)}\right)^*
\end{aligned}
\tag{3.14}
$$

5. *Continuous-time passband to baseband equivalent transformation:*

At the previous step, we ended up with the continuous-time passband formulation of DIDO Volterra. This formulation is used to derive the continuous-time baseband equivalent representation where the envelope of the outputs is related to the envelope of the inputs. The steps of the derivation includes the transformation to the frequency domain by applying a multi-dimensional Laplace transform, applying a frequency translation from the carrier frequency to baseband and going back to time domain. The details of the transformation are available in [26]. The continuous-time baseband expression relates the envelope of the outputs to the inputs, without including the carrier frequency phasor in the basis functions.

The corresponding continuous-time baseband 3rd order response derived from (3.13) is displayed in equation (3.15).

$$\begin{aligned}\hat{y}_{1,3}(t) = &\iiint_{-\infty}^{+\infty} \hat{h}_{3,0}(\tau_1,\tau_2,\tau_3)\ \widehat{BF}_{3,0}(\hat{x}_1(t),\hat{x}_2(t),\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3 \\ &+ \iiint_{-\infty}^{+\infty} \hat{h}_{0,3}(\tau_1,\tau_2,\tau_3)\ \widehat{BF}_{0,3}(\hat{x}_1,\hat{x}_2,\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3 \\ &+ \iiint_{-\infty}^{+\infty} \hat{h}_{1,2}(\tau_1;\tau_2,\tau_3)\ \widehat{BF}_{1,2}(\hat{x}_1,\hat{x}_2,\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3 \\ &+ \iiint_{-\infty}^{+\infty} \hat{h}_{2,1}(\tau_1,\tau_2;\tau_3)\ \widehat{BF}_{2,1}(\hat{x}_1,\hat{x}_2,\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3\end{aligned} \tag{3.15}$$

And,

$$\begin{aligned}\widehat{BF}_{1,2}(\hat{x}_1,\hat{x}_2,\tau_1,\tau_2,\tau_3) = &\big(\hat{x}_1(t-\tau_1)\big)^*\big(\hat{x}_2(t-\tau_2)\big)\big(\hat{x}_2(t-\tau_3)\big) \\ &+ \big(\hat{x}_1(t-\tau_1)\big)\big(\hat{x}_2(t-\tau_2)\big)^*\big(\hat{x}_2(t-\tau_3)\big) \\ &+ \big(\hat{x}_1(t-\tau_1)\big)\big(\hat{x}_2(t-\tau_2)\big)\big(\hat{x}_2(t-\tau_3)\big)^*\end{aligned} \tag{3.16}$$

6. <u>Baseband equivalent Volterra series discretization:</u>

The final step consists of the discretization the continuous-time baseband equivalent Volterra reached previously to be implementable in CAD environment. The discretization takes into consideration the causality of the system, the fading memory assumption and the symmetry with respect to each input separately. Digitizing the 3rd order response yields the following result in (3.17)

$$\begin{aligned}&\hat{y}_{1,3}(t) \\ &= \sum_{i_1=0}^{M}\sum_{i_2=i_1}^{M}\sum_{i_3=i_2}^{M} \hat{h}_{3,0}(i_1,i_2,i_3)\,\widehat{BF}_{3,0}(\hat{x}_1(n),\hat{x}_2(n),i_1,i_2,i_3) \\ &+ \sum_{i_1=0}^{M}\sum_{i_2=i_1}^{M}\sum_{i_3=i_2}^{M} \hat{h}_{0,3}(i_1,i_2,i_3)\,\widehat{BF}_{0,3}(\hat{x}_1(n),\hat{x}_2(n),i_1,i_2,i_3) \\ &+ \sum_{i_1=0}^{M}\sum_{i_2=0}^{M}\sum_{i_3=i_2}^{M} \hat{h}_{1,2}(i_1,i_2,i_3)\,\widehat{BF}_{1,2}(\hat{x}_1(n),\hat{x}_2(n),i_1,i_2,i_3) \\ &+ \sum_{i_1=0}^{M}\sum_{i_2=i_1}^{M}\sum_{i_3=0}^{M} \hat{h}_{2,1}(i_1,i_2,i_3)\,\widehat{BF}_{2,1}(\hat{x}_1(n),\hat{x}_2(n),i_1,i_2,i_3)\end{aligned} \tag{3.17}$$

Where:

$$\begin{aligned}&\widehat{BF}_{1,2}(\hat{x}_1(n),\hat{x}_2(n),i_1,i_2,i_3) \\ &\qquad = \big(\hat{x}_1(n-i_1)\big)^*\big(\hat{x}_2(n-i_2)\big)\big(\hat{x}_2(n-i_3)\big)\end{aligned} \tag{3.18}$$

As a conclusion, the use of this method results in a significant reduction in the number of terms in the digitized baseband formulation of the DIDO Volterra model compared to the direct digitization from the continuous-time DIDO Volterra model. The model extracted is used in simulation to illustrate its capabilities to capture the behaviour of nonlinear devices in a matched environment.

3.2 Simulation Based Model Validation

3.2.1 Envelope Simulation Overview

Simulation of nonlinear circuits and devices is a more complex problem than linear systems. There are several methods available in the literature and implemented in CAD commercial simulators in both frequency domain and time domain. The goal is to build a nonlinear simulator that guarantee an acceptable accuracy and a manageable simulation time. The most commonly used simulation techniques for high-frequency nonlinear circuits are harmonic balance, time domain integration as well as envelope simulation.

Time domain integration techniques [27] solve partial differential equations using different integration methods and are able to reproduce the transient behavior of the simulated circuit. Traditional time domain techniques struggle to find the steady-state response of the circuits since they have to simulate the circuits long enough for the transient behavior to vanish. Alternatively, harmonic balance [28] is a frequency domain technique that provides the steady-state response of a circuit at fundamental frequencies, mixing products and harmonics. It uses a system of algebraic equations to represent the circuits instead of the differential equations. There are different ways to solve these equations, among them the Newton- Raphson method is the most used.

As far as the nature of the signal is concerned, harmonic balance is used with periodic and pseudo-periodic signals to simulate mild nonlinearities, but when used to simulate strong nonlinear circuits with modulated signal, it suffers from dramatic increase in complexity and may not converge. On the other hand, time domain integration techniques are suitable to the slow-varying baseband signals and are very slow when used with modulated signal around high-frequency carrier. Hence, envelope simulation [29] were developed to be a compromise between the two previous techniques. Envelope Simulators represent the modulated signals as discrete frequency components with time-dependent modulation, *i.e. envelope signal*, around them. The signal envelope is handled by time-domain integration techniques, while harmonic balance handled the high-frequency carrier allowing both for transient and steady-state response of the circuits to be captured.

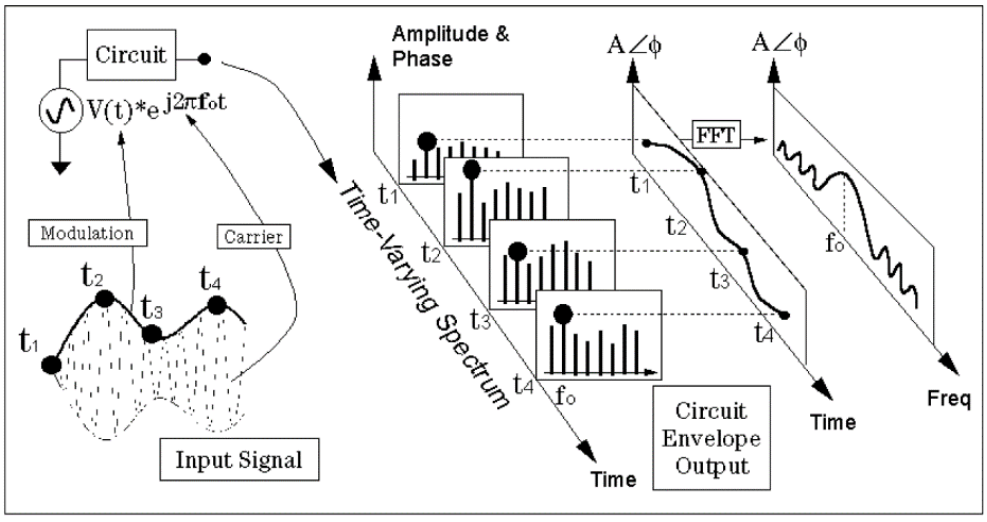


Figure 3.3 - Envelope Simulation Process [30]

Figure 3.3 describes the envelope simulation schematic in the Agilent Advanced Design System (Agilent ADS), the commercial software used to collect the measurement data needed to extract the DIDO Volterra model. Each modulated signal is represented by an envelope that modulates a carrier. The complex (amplitude and phase) samples of the envelope are computed

and for each time step, a harmonic balance analysis is performed. This procedure characterizes the time-varying envelope through a time-varying spectrum. Since the envelope is sampled instead of the RF signal, the simulation bandwidth must be equal to the bandwidth of the modulation signal, which is significantly lower than the carrier frequency.

3.2.2 Simulation Schematic

The simulated DUT can be any two-port network. It can be passive or active, linear or nonlinear microwave device. Examples are analog filters, transistors and RF power amplifiers. As long as a circuit representation or a compact model of the device is available, the behaviour model could be extracted and compared to the circuit results. Since compact models reproduce similar trends than an actual DUT when measured in a characterization setup, the simulation results are accurate enough to validate the model. The simulation based validation of the DIDO Volterra model is chosen to separate the potential sources of errors, namely model related or measurement setup related errors.

The schematic used to simulate the nonlinear device in an unmatched environment is described in this sub-section and its important features are outlined. The simulation architecture is developed to be similar to the measurement setup (see Figure 3.5). Indeed, the schematic mimics an active load-pull setup by using DC block, couplers as well as circulators. The idea behind including such components is to enable the possibility of representing these components by their equivalent S-parameters in a future work. The schematic can be divided in three parts: the modulated signal generation output, signal measurement and load impedance control (active load-pull).

Signal generation is achieved by loading an I/Q file that contains the time domain samples of the baseband signal and use it to modulate the carrier frequency. I/Q samples could be synthesized using any CAD software and can represent any type of modern modulated signal (LTE, WCDMA …). Figure 3.4 presents the schematic used to generate the signal for each source.

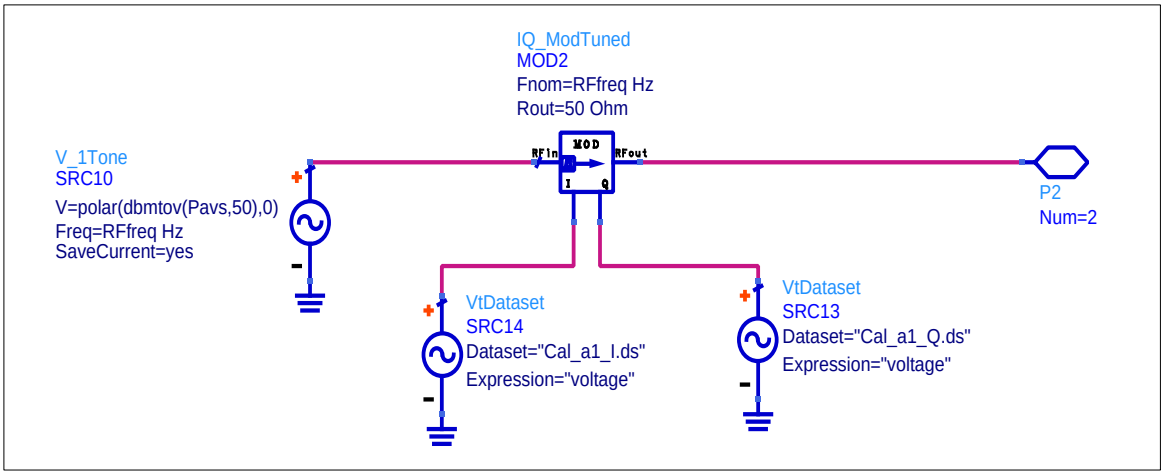


Figure 3.4 - Modulated Signal Generation

Two sources are included in the schematic: one source to provide the main excitation signal, $x_1(t)$, at the input port of the DUT while the second source is needed to synthesize the second source, $x_2(t)$, to perform the active load-pull by emulating a variation in the load impedance.

Signal measurement is done using two directional couplers to separate incident and reflected power waves. Although the circuit representation in the simulator is in terms of voltages and currents, the signals are transformed to power waves representation using the linear equations 2.1 and 2.2. This transformation is done by including ABCD matrices that follow the couplers in order to mimic the measurement system by adopting the same representation. Finally, the resulting power waves are demodulated and the envelopes of the various signals are extracted in both ports of the DUT.

The load control is achieved using an active load-pull technique. A circulator is used to redirect the output power of the DUT to a matched load where it is absorbed and actively feed another signal from the second source in order to emulate a user-defined load. The choice behind implementing active load-pull, is the ability to inject an independent signal at the second port, independently from the first port. In fact, if a_2, the incident wave at the second port of the DUT, is merely the reflection of b_2 with a certain reflection coefficient (Γ_L) while the model is trying to find a function that model b_2 with the respect to a_1 and a_2 $\left(b_2 = fct\,(a_1, a_2)\right)$; the model will totally ignore a_1 and model b_2 as $b_2 = \Gamma_L \,.\, a_2$.

Another important aspect that should not be ignored during the simulation is the harmonic terminations of the DUT. Although the impedance control is realized only at the fundamental frequency, the harmonic terminations have to be set to specific values. Different harmonic terminations will result in different performance. In measurements, the same harmonic terminations must be used. For the case of a RF power amplifier, different harmonic terminations are associated with different classes of operation. This will dictate the waveforms shape in the time domain and will affect the RF power amplifier performances in terms of efficiency and output power.

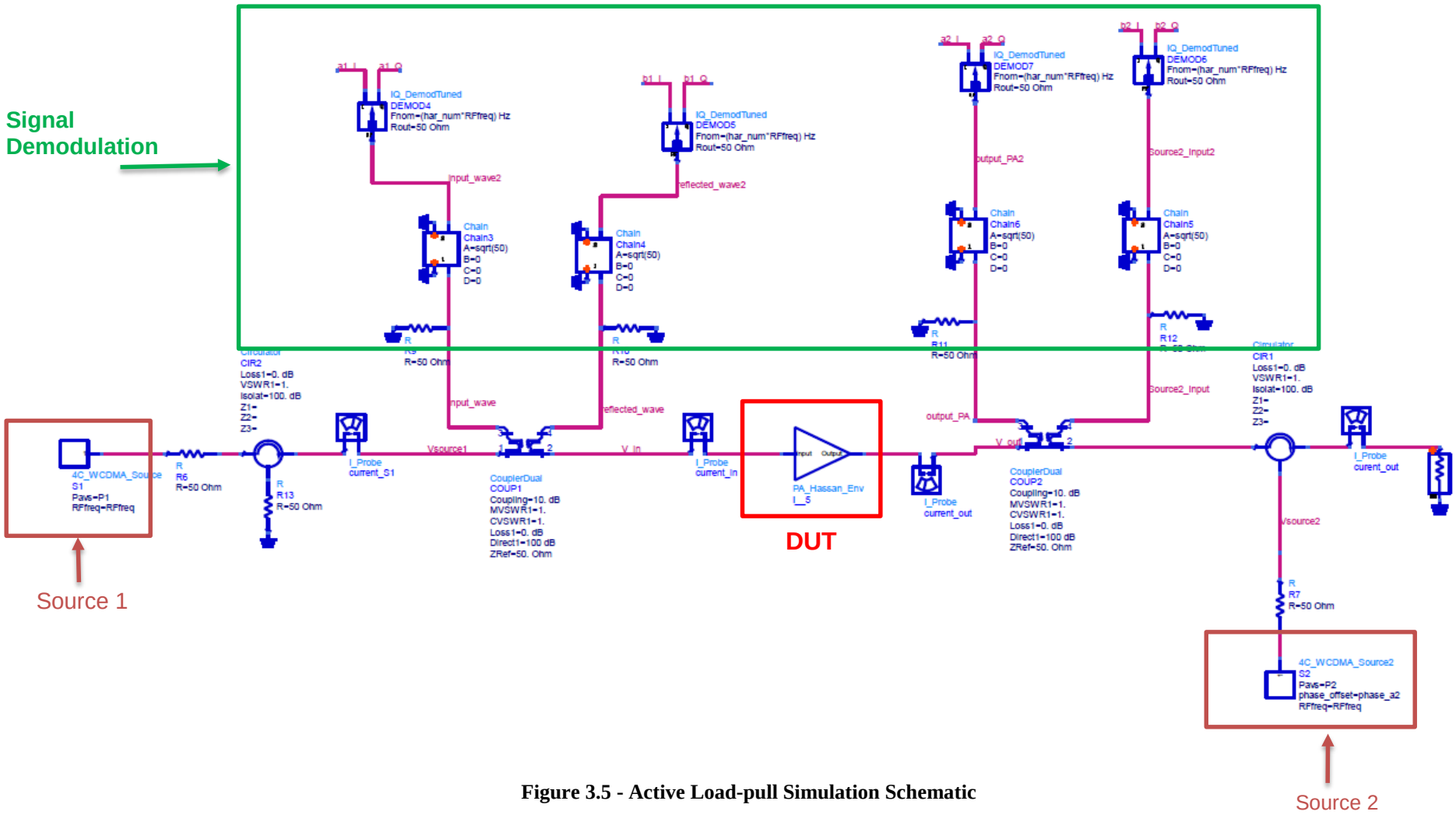


Figure 3.5 - Active Load-pull Simulation Schematic

3.2.3 Model Extraction Procedure

In order to extract the DIDO Volterra model coefficients, the procedure described in Figure 3.6 is followed.

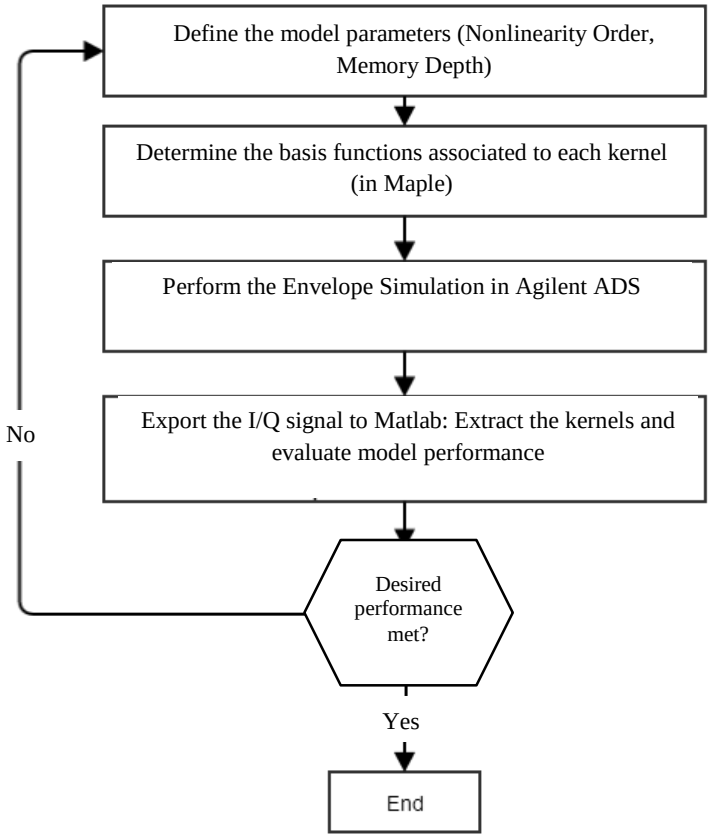


Figure 3.6- Extracting Volterra DIDO Model

First, the parameters of the DIDO Volterra model, namely N and M, are set. This step is about the choice of three parameters which are the total order of nonlinearity N and the memory depth of each input M_1 and M_2. The initial values of these three parameters relies on the experience of the user and could be tuned later based on the simulation results to find the better compromise between complexity and accuracy.

Secondly, the process of determining the *basis functions* is accomplished using the method explained before for complexity reduction. The basis functions being the combination of the input instants at different instant and different nonlinearity order associated with each kernel. Once, this task is performed it will enable the model to be reformulated as a linear problem with respect to its kernels and for LSE algorithm to be used to extract the kernels.

Next, the simulation of the DUT, which can be any nonlinear circuit, is performed using the Agilent ADS schematic detailed above. The user-defined modulated signal is applied to DUT at the appropriate frequency and average power level. The load impedance is varied around a certain value using the active signal from the second source. The envelope signal (I/Q samples) of both the incident and reflected power waves at the port of the DUT is saved into a text file so they can be processed at a later stage.

Once the data from the simulation is available, it is used as input to a Matlab script which executes the extraction of the model coefficients using LSE algorithm and compares the model

performance to the simulation results. The coefficients are written in a matrix formulation (equation 3.19).

$$\underbrace{\begin{bmatrix} bf_{11} & \cdots & bf_{1n} \\ \vdots & \ddots & \vdots \\ bf_{n1} & \cdots & bf_{nn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}}_{b} \tag{3.19}$$

where A is matrix built from the basis functions, X is the vector of kernels and b is the output vector.

The model performance is evaluated using metrics in both time domain and frequency domain. In time domain, the NMSE given in equation 3.9 is calculated. NMSE measures the deviation of the model samples from the simulation samples in average. It is usually reported in logarithmic scale (dB) and the lower the value is, the better the model. In frequency domain, the spectrum of the model is obtained through the Fourier transform and compared to the spectrum from the simulation.

$$NMSE = 10 * log_{10}\left(\sum_{n}\left|\frac{\left(\left(y(n)-\hat{y}(n)\right)^2\right)}{\sum_n|\hat{y}(n)|^2}\right|\right) \tag{3.20}$$

where $y(n)$ is the simulation reference signal and $\hat{y}(n)$ is the model signal.

Finally, if the model performance is not satisfactory, the model parameters are changed and the extraction procedure is redone. Also, the simulation conditions could be swept in terms of input average power level, load reflection coefficient, biasing conditions, etc. The model coefficients are extracted for each case and tabulated to build a comprehensive model.

3.2.4 Simulation Based Model Validation

The DIDO Volterra Model was tested on an RF power amplifier to assess its performance. The DUT is the ADS schematic of 45 Watt GaN PA internally designed in the research group, fabricated and validated through measurements. Hence, in a future step the model will be extracted from measurements. The RF PA circuit representation is based on source/matching networks build around the Cree packaged transistor model that has already showed good agreement with experimental results. A comparison will be established between the predicted behaviour of the DIDO Volterra model and PA circuit representation that uses the large-signal model provided by the manufacturer.

The operating conditions under which the simulation were conducted are described in Table 3.2. It is important to note that the harmonic terminations is an important criteria that should be included to define the simulation conditions with a tendency to be overlooked.

Table 3.2 - Simulation Conditions

Simulation Conditions	Values
Centre Frequency	850 MHz
Stimulus	1C - WCDMA signal
DUT	45 Watt GaN RF PA
Average Input Power	30 dBm
Harmonic Terminations	Matched Harmonics
Simulation type	Envelope Simulation

The DIDO Volterra model to be extracted here has a total nonlinearity order equal to 3 and memory depths equal to 2 ($N = 3$; $M_1 = 2$, $M_2 = 2$). The model NMSE performance of different matching conditions is highlighted in Table 3.3.

Table 3.3 - NMSE Performance of different Matching Conditions (3th order Model)

Mismatch Conditions	Matched (a_2=0)	Low Mismatch	Relatively Low Mismatch	Relatively High Mismatch	High Mismatch	Number of Kernels
DIDO Volterra (N=3; M_1=M_2=2)	-	-32.8	-31	-28.3	-26.3	**24**
SISO Volterra (N=3; M=3)	-35	-	-	-	-	**21**

The matching conditions are set by varying the average power of the waveform a_2 injected at the output port of the device. The higher is the average power of a_2, the higher is the mismatch. When the load is perfectly matched, i.e. a_2 equals to zero, the DIDO Volterra model falls back to the SISO case and there is no difference between the two models. The NMSE reported in the above table is for modeling the output waveform at the DUT port 2 $\left(b_2 = fct(a_1, a_2)\right)$ and similar results are expected when modeling the output waveform at the DUT port 1 $\left(b_1 = fct(a_1, a_2)\right)$. It can be noticed that NMSE of the DIDO Volterra model drops at high mismatch where the reflection coefficient is around 0.9. Indeed, at high mismatch, the high average power of a_2 introduces stronger nonlinearity and higher order model should be used at the expenses of higher complexity. When SISO Volterra model was used to model $b_2 = fct(a_1)$ and ignoring the mismatch caused by the presence of a_2, the resulting NMSE was very low reflecting the limited capability of the SISO model in an unmatched condition.

In addition to 3rd order total nonlinearity model, 5th order model have been studied for comparison purposes. The results are reported in Table 3.4. It can be noticed that there is an improvement in the NMSE values with the increase of nonlinearity order at the expense of a higher number of kernels.

Table 3.4 - NMSE Performance for Different Matching Conditions (5th order Model)

Mismatch Conditions	Low Mismatch	Relatively Low Mismatch	Relatively High Mismatch	High Mismatch	Number of Kernels
DIDO Volterra (N=5; M_1=M_2=2)	-33.6	-32.1	-29.7	-27.8	**80**
DIDO Volterra (N=5; M_1=3, M_2=2)	-35	-34.1	-32.2	-30.4	**166**
DIDO Volterra (N=5; M_1=3, M_2=3)	-35.2	-34.5	-32.6	-30.9	**314**

To further monitor the mismatch extent, the variation of the load impedance resulting from the interaction between the DUT output at port 2 and the injected power from source2, the load impedance was represented in a Smith Chart and reported in Figure 3.7. The area around the center of

the Smith Chart represent the different load impedance presented to the DUT during the model extraction.

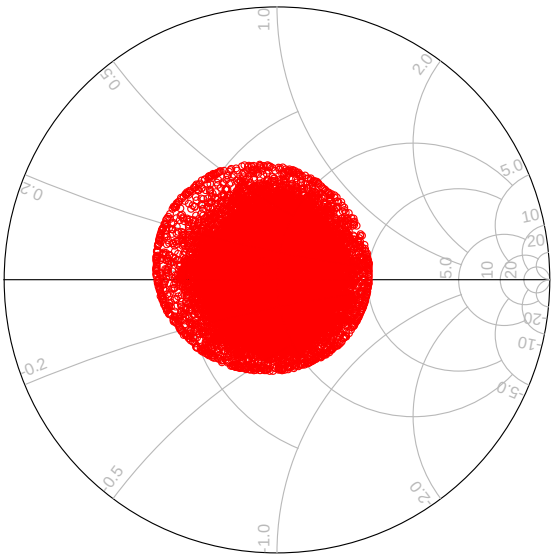
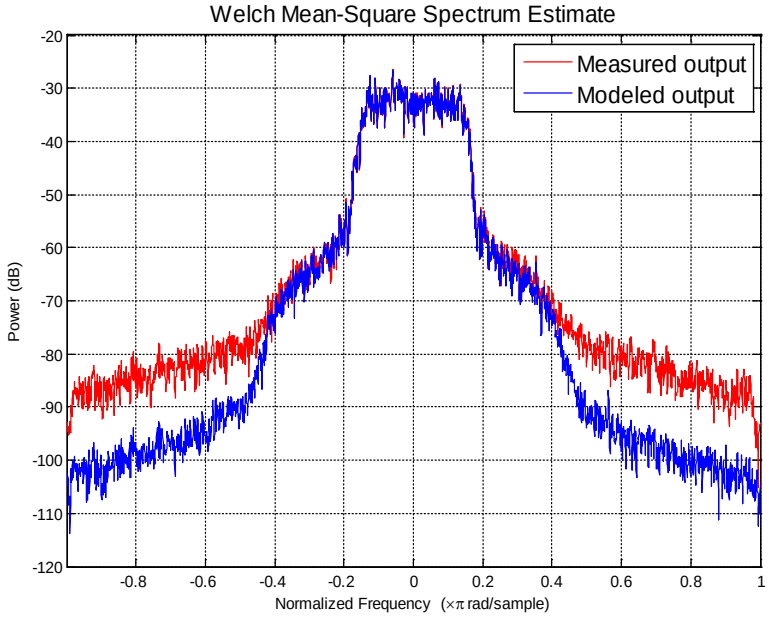
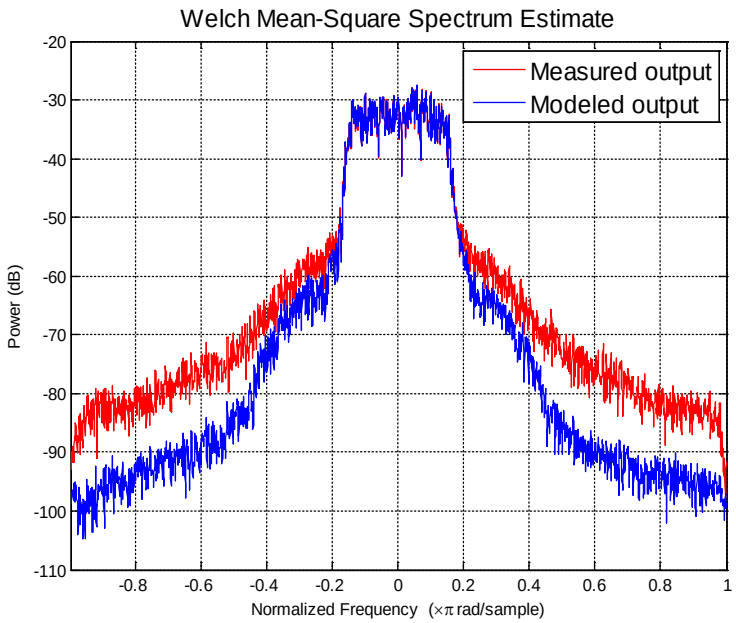


Figure 3.7 - Load Impedance Coverage

In addition to the NMSE, both the spectrum predicted by the model and the output spectrum are compared. Figure 3.8 shows the spectrum of DIDO Volterra model with the 3rd nonlinearity order and memory depths equal to 2 ($N = 3$; $M_1 = 2$, $M_2 = 2$).



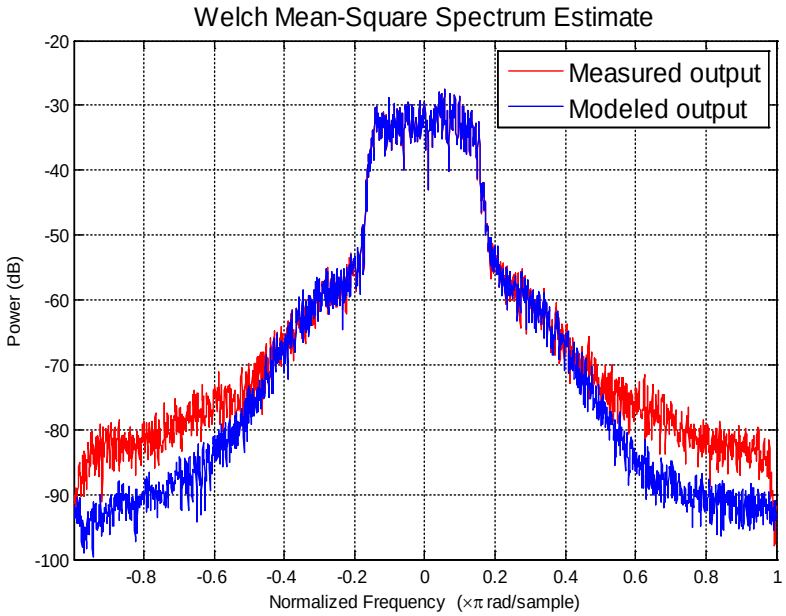
(a) Low Mismatch



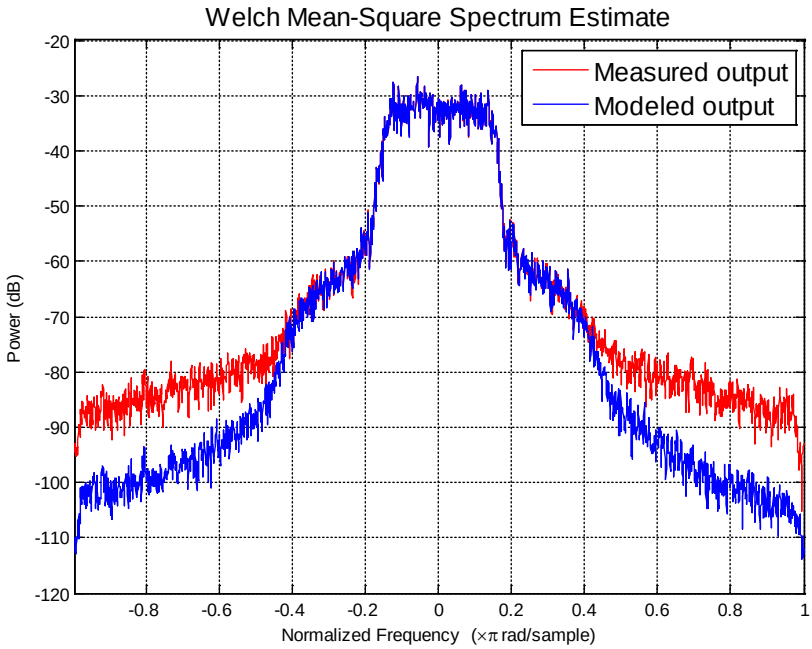
(b) High Mismatch

Figure 3.8 - 3rd Order Model Spectrum (a) Low Mismatch (b) High Mismatch

Based on the spectrum showed in Figure 3.8, it can be concluded that 3rd order nonlinearity model is enough for the case of low mismatch but for higher mismatch, the model suffers from a degradation in its performance. This observation confirms that increasing the area of load coverage, i.e. the possible load impedances that can be presented to the device, dictates higher nonlinearity order for the model. Thus, the model nonlinearity was increased to 5 while keeping the same memory depth ($N = 5$; $M_1 = 2$, $M_2 = 2$). Figure 3.10 presents the corresponding spectrum to nonlinearity order 5 which has better accuracy in both low and high mismatch.



(a) Low Mismatch



(b) High Mismatch

Figure 3.9 - 5rd Order Model Spectrum (a) Low Mismatch (b) High Mismatch

Overall the initial investigation of the DIDO Volterra model demonstrates promising results and shows the model capabilities to capture the nonlinear behaviour of the device. The simulation based validation needs to be confirmed through realistic measurements. The simulation is used as an intermediate step to separate measurement errors from model errors. The methodology to perform such measurement will be clarified in the next chapter.

Chapter 4

Characterization Setup under Modulated Signals

After the validation of the model through simulation, the DIDO Volterra model needs to be extracted and validated using nonlinear measurements data with modulated signals. Realizing a nonlinear characterization system that uses modulated signals is a complicated task that need to take into account many practical difficulties and implementation challenges. Any issues or limitations in the measurement system will affect dramatically the performance of the model. The platform should allow the user to measure the currents and voltages at the DUT ports within an unmatched environment in time domain, the nature domain for the RF circuit design theory. It should increase the accuracy and reliability of measurement-based models when extracted under modulated signals. Moreover, the platform should provide the necessary data to extract and validate DIDO Volterra model or any time domain model.

In this chapter, the challenges as well as the overall guidelines to build a nonlinear characterization platform using modulated signal are detailed. The choices made and the accomplished steps are highlighted with particular focus in the calibration routines that needs to be implemented and the equipment involved in the process.

4.1 Characterization Setup Overview

As explained in Chapter 2, a nonlinear characterization setup for unmatched nonlinear devices is composed of three parts: waveform generation, waveform engineering approaches and waveform measurement instruments. Depending on the purpose of the characterization setups as well as the available equipment, choices have been made in each of these three categories. Figure 4.1 presents the different parts that constitute the proposed characterization setup building block and connections.

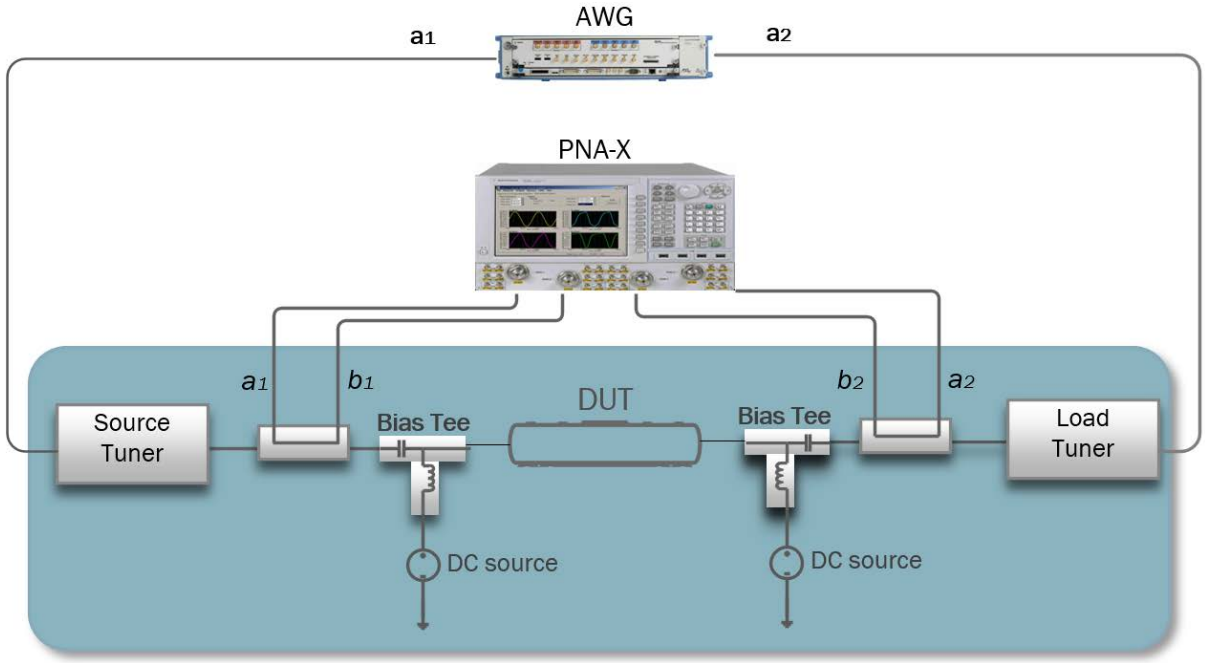


Figure 4.1- Characterization Setup Block Diagram

The waveform generation: The waveform generation function should be able to produce a modulated signal with the required bandwidth and purity. The signal purity is measured generally in terms of spurious free dynamic range (SFDR) and adjacent channel power ratio (ACPR). These two metrics characterize the generated signal's power with respect to the spurious and undesired frequency components power. Modern setups use high speed AWGs for wideband signal generation. High speed AWGs take the digital baseband samples of a modulated signal (I/Q) and produce the signal's analog envelope at the desired intermediate frequency (IF). The signal can be up-converted in a latter step from IF to the carrier frequency of interest, although the modern AWGs' IF reaches few GHz which is enough for most applications.

AWGs are built around DACs, thus their specifications includes mainly the sampling frequency, bit resolution, analog bandwidth and memory size. The available AWG to conduct this work was the Agilent AWG M8190A [31] with 12-bit resolution, 12 GSa/s sampling frequency, 5 GHz analog bandwidth and 90 dBc SFDR which is enough for the purpose of this experiment. Furthermore, the 5 GHz analog bandwidth is wide enough to accommodate modulated signals at baseband, fundamental frequency and the harmonics, given that the fundamental frequency is not too high. For example, if the fundamental frequency is chosen to be around 1 GHz, modulated signal can be generated up to the 4^{th} harmonic as it is still within the analog bandwidth.

The waveform engineering: Waveform engineering, in its broad sense, is the ability to control the voltages and currents at the DUT ports. This is achieved by setting and measuring the source/load impedances at the DUT ports, known as source/load pull systems. A load-pull system, as detailed in Chapter 2, can be passive, active with different possible configuration or hybrid.

Since the DIDO Volterra model to be extracted is an envelope domain model, the natural choice would be a load-pull system that works with signal envelopes. Using the signal envelope also avoids the instability issues usually associated with active load-pull and allows to shape the response of the reflection coefficient in order to mimic real life scenarios. For example, if the DUT is an RF power amplifier, the reflection coefficient could be changed versus frequency to track the frequency dispersive impedance of a varying antenna at the output port of RF power amplifier.

Once the stability issue is solved, an active load-pull is more beneficial to using passive tuners. Indeed, it has a wider bandwidth, and overcomes the losses problem of passive tuners and it can present a load reflection coefficient that reaches one, which is equivalent to fully reflecting the output travelling wave. However, on top of the active load-pull, passive tuners could be added to the setup for the purpose of maximizing the power transfer. If an RF transistor is to be modeled, tuners should be used to conjugate match the input and output impedance of the transistor in order to guarantee maximum power transfer. Failing to do so will force the usage of high power drivers in order to compensate for the lower transferred power and attain the necessary power level. Special care must be given to the tuners' frequency response which should not distort the modulated signal stimulus.

The waveform measurement instruments: They are measurement systems able to measure power waves in time or frequency domain in order to reconstruct the voltages and currents at the DUT ports. This ability to measure correctly the power waves does not imply necessarily the ability to set or control them as is the case in waveform engineering. The different possible waveform measurement instruments are high speed oscilloscopes, subsampling-systems like LSNA and mixer-based systems like NVNA. The choice of the waveform measurement instrument to be used is the NVNA, specifically Agilent PNA-X [14]. Indeed, the NVNA will give the flexibility to use CW, pulsed CW in top of the modulated signals. More importantly, the presence of the phase reference capacity that guarantees the relative phase coherence between fundamental and harmonics, is an important feature that could be used in future extension of the DIDO Volterra from fundamental frequency to

harmonics. Additionally, NVNA provides a dynamic range that can be superior to 100 dB which is very important to ensure accuracy measurement of power waves.

The Agilent PNA-X is meant to use CW signals as stimulus. It is based on superheterodyne architecture: the signal at the carrier frequency is down-converted to an intermediate frequency where the appropriate signal conditioning (filtering, gain, attenuation …) is applied before being digitized. The waves are measured using the frequency swept-mode where each frequency component is captured at a time. The coherency between the different receivers is achieved by ensuring the fact that they all share the same LO signal. Figure 4.2 shows the main components that constitute the PNA-X. Although, the PNA-X has two internal sources, they are not used in the setup in this work as they only generate CW signals and an AWG is used instead, therefore the PNA-X is treated as a 8-channels phase coherent receiver.

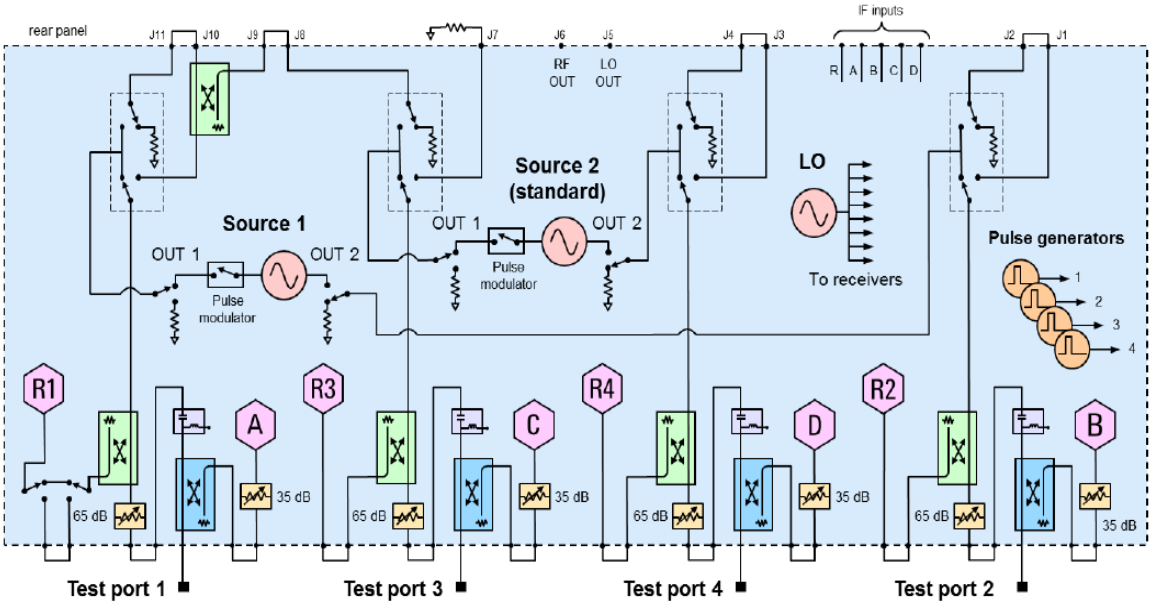


Figure 4.2 - Agilent PNA-X N5242A Block Diagram

In order to measure modulated signals, the frequency swept-mode is not convenient as capturing the frequency components one by one within the signal bandwidth extends dramatically the measurement time. Also, the measured frequency components have to be phase coherent which is difficult to realize especially in low frequency resolution. Thus, the frequency swept-mode is replaced by a time domain capture.

In a time domain capture, the row time domain samples are stored at the output of the analog to digital convertor (ADC) (see Figure 4.3) where any filtering in the IF path that could limit the signal bandwidth is bypassed. The stored signal is then demodulated in a post-processing step and the envelopes are extracted. The disadvantage of applying time domain capture, is the reduction in the dynamic range due to the usage of wide IF bandwidth. To enhance the dynamic range, coherent averaging of the time domain samples and taking long time sequence has been used to reduce the noise floor in depends of a longer measurement time. In conclusion, in order to apply time domain capture, a tradeoff between measurement time and dynamic range must be reached.

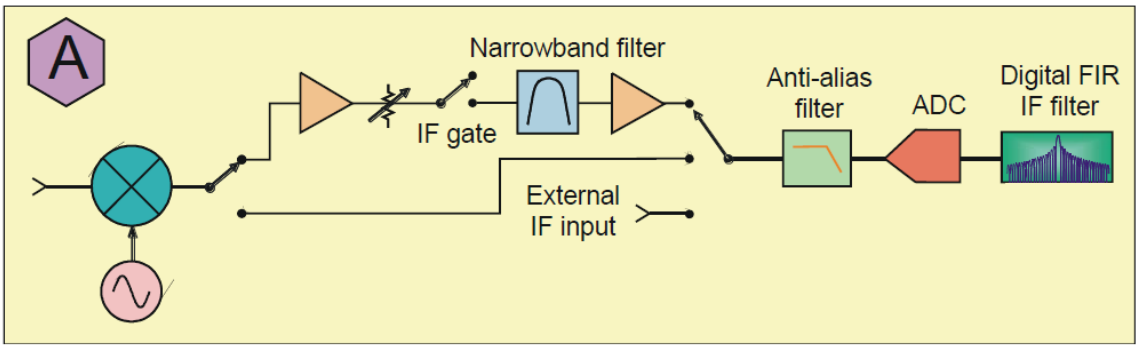


Figure 4.3 - Agilent PNA-X Receiver Block Diagram

4.2 Setup Calibration Routines

Traditionally, network analyzers are used to measure S-parameters which are a rationed measurement between the waves travelling in the forward and reverse directions. Rationed measurements only need a vector calibration to be performed which should de-embed the imperfections of the instrument. Also, it allows to set the reference plans right at the ports of the DUT and to de-embed any component in the path of the signal such as linear driver, attenuator or fixtures. In the measurement science, the theory to perform such calibration has matured enough and different algorithms has been developed to accommodate several practical situations (high frequency, on-wafer measurements …) [32].

However, in this work, the NVNA is expected to make unrationed measurements of each travelling wave separately (a_1, a_2, b_1, b_2). For this purpose, on top of the vector calibration, a power calibration and a phase calibration are needed. The power calibration uses a power meter because of its traceable and reliable precision, while the phase calibration establishes the phase coherency between fundamental and harmonics of the same wave.

All of these three types of calibrations can be grouped under the category of RF calibration that corrects for the errors between the receivers and the DUT. If the signal is a CW, this calibration is enough, but since the signal is a modulated signal with wide bandwidth, the IF response cannot be ignored and the frequency response of the receiver IF path, just after the mixer, needs to be corrected.

4.2.1 Receiver IF Calibration

The IF calibration characterizes the frequency response of the receivers and compensates for its magnitude and phase response. If the signal is narrowband, this calibration can be avoided; but with wideband modulates signals, it is mandatory step. The reason is that once down-converted to the IF frequency, the signal occupies a wide-bandwidth and is subject to frequency dispersions of the IF circuit response. With the Agilent PNA-X receivers used to measure the signals, it should be noted that the travelling waves are measured using the time domain capture where the row samples at the output of the receiver ADC are stored, demodulated and post-processed to extract the envelope data.

The IF calibration is performed on the signals' envelopes and not on the RF signals around the carrier frequency. As an illustration, the row response of a PNA-X receiver is displayed in Figure 4.4 where it shows almost 11 dB roll-off. The calibration procedure is repeated for each one of the PNA-X receivers. If the signal reaches the receiver without distortions and the IF calibration is applied correctly, the receiver reading after correction should be identical to the signal at its input.

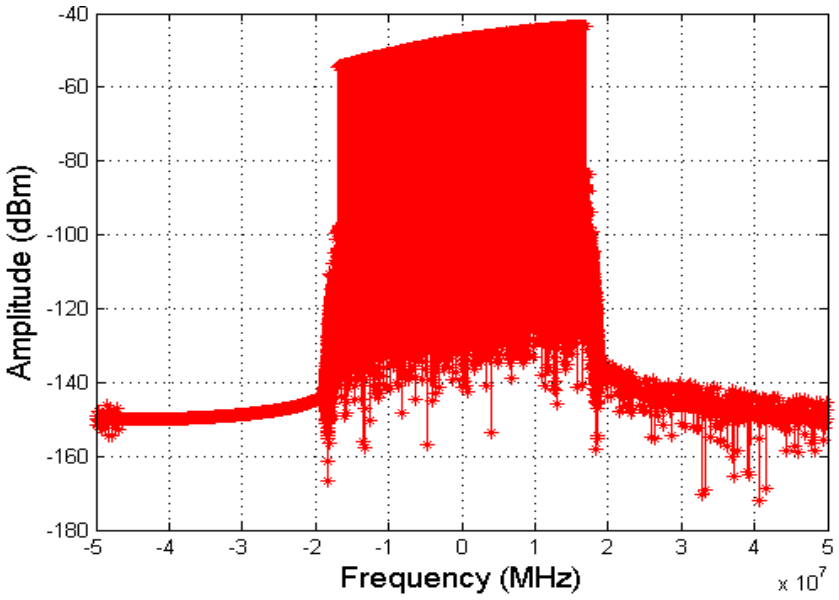


Figure 4.4 – Raw Baseband Spectrum of PNA-X Receiver C

The first step in the calibration is to generate a multitone signal using the AWG. There is no general rule for selecting the number of tones to use, since it depends on the smoothness of the receiver frequency response, but several iterations could be performed with increasing number of tones until acceptable results are reached. Then, the AWG is connected directly to one receiver of the PNA-X as shown in Figure 4.5. The PNA-X and the AWG must share the same 10 MHZ reference to exclude any frequency offset between the two instruments. Moreover, triggering is necessary in order to time align the signal at the output of the AWG and the measured signal. The calibration should cover the measurement bandwidth which is in general 3 to 5 times the signal bandwidth. The maximum measurement bandwidth is limited by the sampling rate of the ADC.

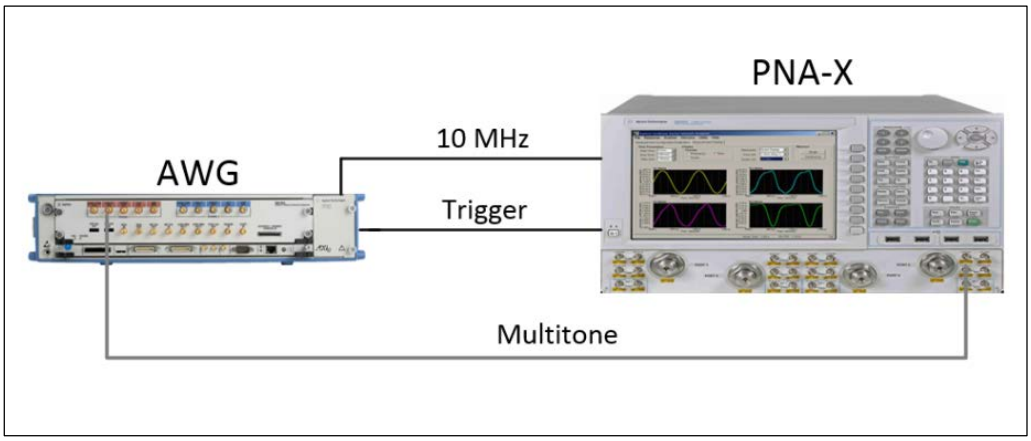
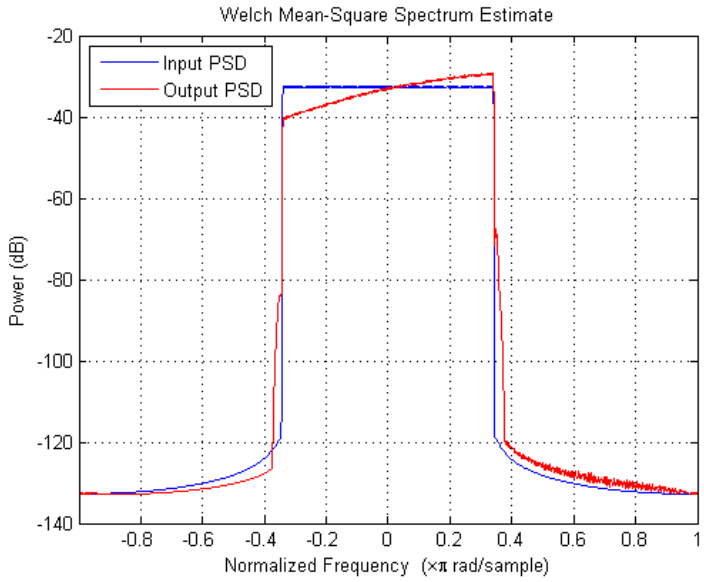


Figure 4.5 - IF Calibration Setup

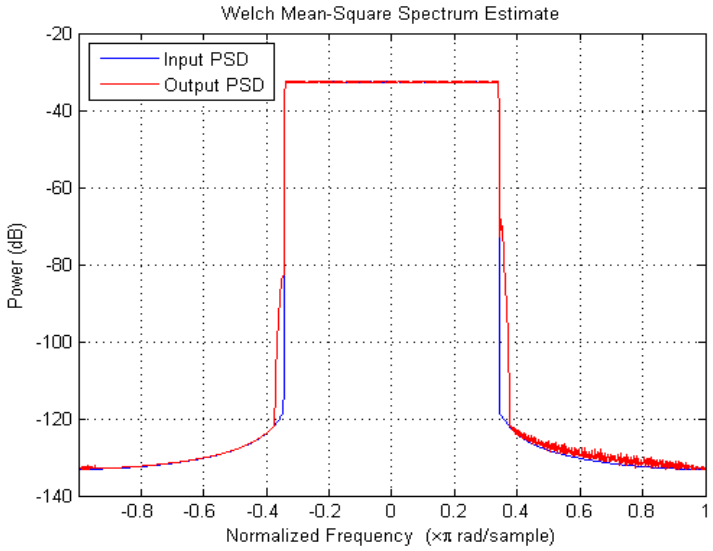
The next step to data acquisition is to compute the frequency spectrum of the signal envelope which yields a spectrum similar to figure 4.4. The IF frequency response of the receiver is captured using the receiver reading and the signal envelope sent from the AWG. This frequency response is supposed to characterize the receiver and should be constant from one measurement to another if the receivers are not driven in nonlinear regime. IF correction consists of transforming the envelope to frequency domain, using the inverse of the IF frequency response to correct for the distortions due to the receiver and converting the signal back to time domain. An important assumption necessary for this approach is that the signal at the AWG output is distortions free.

Without this assumption, it would not be possible to separate the receiver distortions from the distortions due to the AWG. The assumption on the AWG output signal purity was validated separately in different experiment with a calibrated spectrum analyzer.

Figure 4.6 shows the spectrum before and after correction. The bleu spectrum represent the signal at the output at the AWG and the red spectrum is the receiver output. The measurement was made for 5 MHz 1C-WCDMA signal.



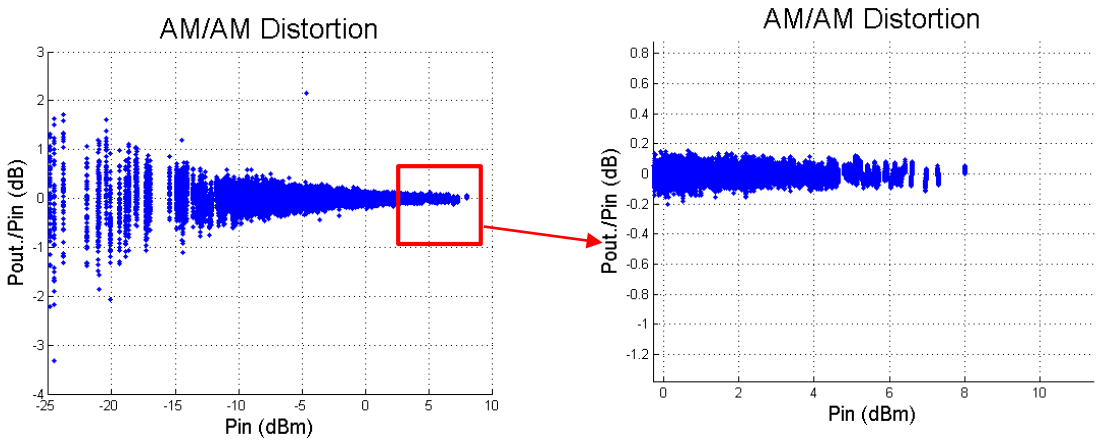
(a) Receiver Response before Correction



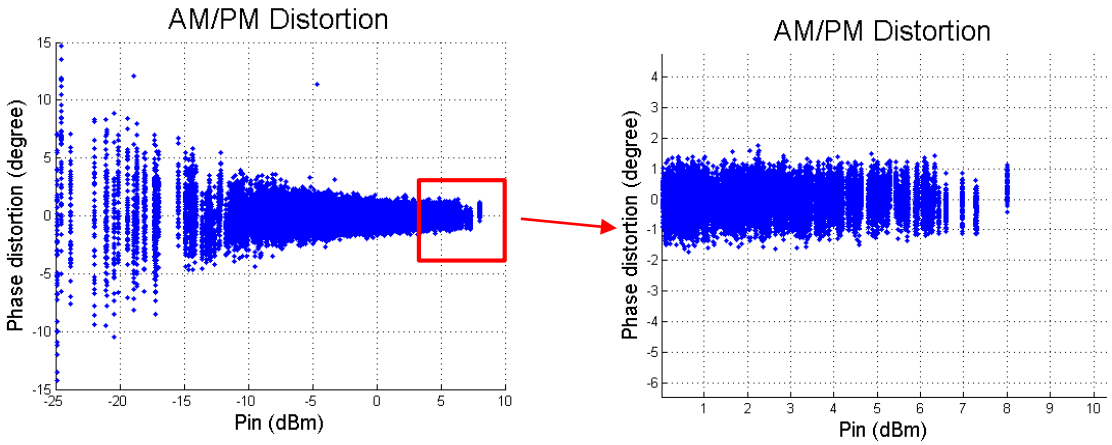
(b) Receiver Response after Correction

Figure 4.6 - Receiver Response (a) Receiver Response before Correction (b) Receiver Response after Correction

Figure 4.7 shows IF calibration validation through AM/AM and AM/PM diagram. For a back-off 15 dB and after the correction, the AM/AM distortion is under 0.2 dB and AM/AM distortion is under 2 degrees.



(a) AM/AM Diagram



(b) AM/PM Diagram

Figure 4.7 Receiver Characteristics (a) AM/AM Diagram (b) AM/PM Diagram

4.2.2 Receiver RF Calibration

4.2.2.1 Vector Calibration

There are three types of errors in a measurement setup: systematic errors, random errors and drift errors. Systematic errors are due to the setup imperfection. They are time-invariant and consequently predictable errors that can be characterized, modeled and mathematically corrected. On the other hand, random errors are time-variant and are mainly related to the noise in the measurement system. They cannot be corrected for through calibration since they are unpredictable and they affect the repeatability of the measurements. Howerver, averaging is generally carried out using different measurements to reduce their effects as well as an increase in the signal power to achieve more dynamic range. Finally, the drift errors represent a change in the setup proprieties, especially when there is a variation in the temperature. Drift errors limit the calibration validity period and often force the user to redo it unless further calibration terms are used to account for drift effects.

The vector calibration in question considers only the systemic errors. There are six possible sources for systematic errors related to leakage, signal reflections as well as impairment in the signal paths. Figure 4.8 gives an illustration of these six errors which are directivity, source mismatch, load mismatch, reflection tracking, transmission tracking and cross-tracking [33].

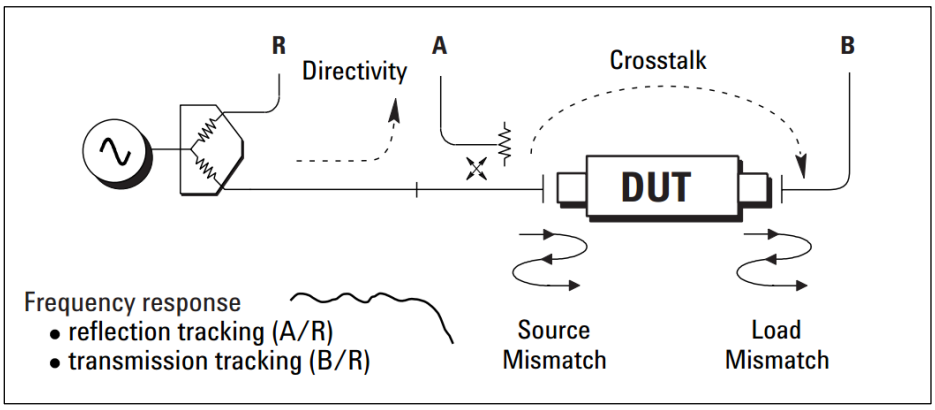


Figure 4.8 - Systemic Error Sources [33]

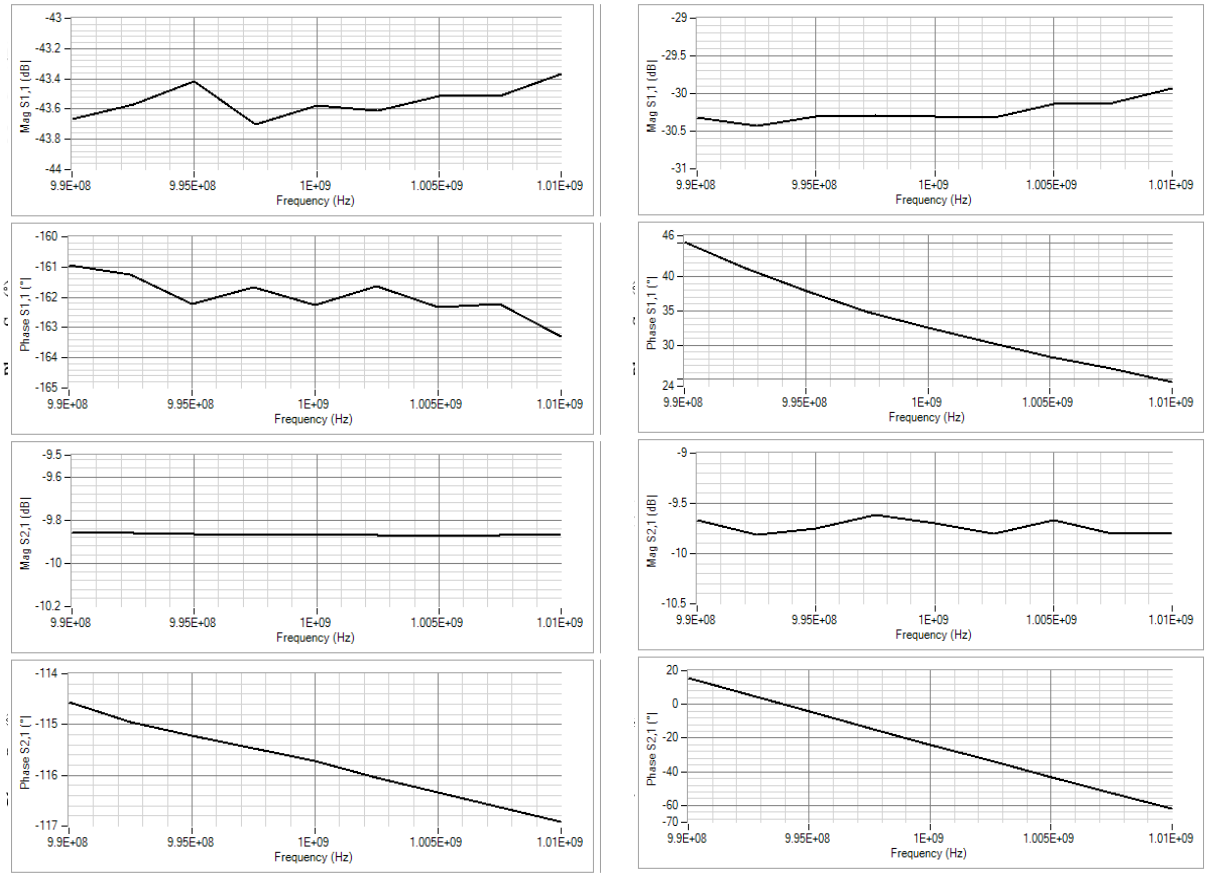
To perform a two-port rationed measurement such as S-parameters, the source and load need to switch from when port to another. Hence, adding the six error terms in the forward and the reverse measurement directions results in a 12-term model that fully characterizes a two-port network. However, the literature contains several derivations and similar error like the popular simplified 8-term error model which is the result of ignoring the switching terms that account for the reconfiguration of the system when measuring a two-port DUT. In order to identify any of these error models, known mechanical or electrical standards are used which should provide linearly independent equations equal to the number of model terms.

Two of the most popular methods used to identify the error terms are Short-Open-Load-Thru (SOLT) and Thru-Reflection-Line (TRL). As the name suggests, SOLT uses different terminations to extract the error terms. First, short, open and load are applied at each port of the equipment. Then, the two ports are connected with a through. If the cross terms are to be included, a load is connected to each port and the transmission coefficients are measured in both directions to characterize the ports coupling. TRL uses a similar procedure to SOLT for error terms identification but does not generally include a load, instead a transmission line is used. TRL is generally used with waveguides or probe stations to perform on-wafer measurements. It can be more accurate that SOLT in a load-pull system especially when used with high frequency because it is more difficult to synthesize and guarantee the performance of a load.

For the purpose of this application, SOLT calibration is used. For generalizing SOLT calibration to modulated signal, a resolution Frequency (F_{res}) is first defined by setting the length of envelope data sequence (N) in the AWG and its sampling frequency (F_s) through equation (4.1). For example, for a 20 MHz measurement bandwidth sampled at 100 MSa/sec, a frequency resolution of 20 KHz will necessitate the use of 20,000 samples.

$$F_{res} = \frac{F_s}{N} \tag{4.1}$$

For each of these frequency component, the error terms are extracted and the correction is performed in the frequency domain. It should be noted that the IF calibration should enabled when extracting and applying the error terms. As a validation procedure, row S-parameters and corrected S-parameters of a 10 dB attenuator were compared. *Figure 4.9* shows phase and amplitude of S_{11} and S_{21} (S_{12} and S_{22} yield similar results). These results were obtained by interfacing Matlab with METAS VNA Tool II software [34].



(a) Row S-parameters (b) Corrected S-parameters

Figure 4.9 - 10-dB Attenuator S-parameters (a) Row S-parameters

(b) Corrected S-parameters

4.2.2.2 Power Calibration

The vector calibration can assess the power ratio between two receivers, but is not able determine the absolute power of each wave. For absolute power reading to be correct, a power calibration needs to be performed. The DUT is replaced by a power meter and is connected to the receiver. The power meter provides the correct power level at the DUT reference plan which will be used to correct the offset in the receiver reading.

Power calibration for two-port network needs only to be conducted at one port since the remaining port power could be deduced from the already performed vector calibration. *Figure 4.7* shows the connection that needs to be made.

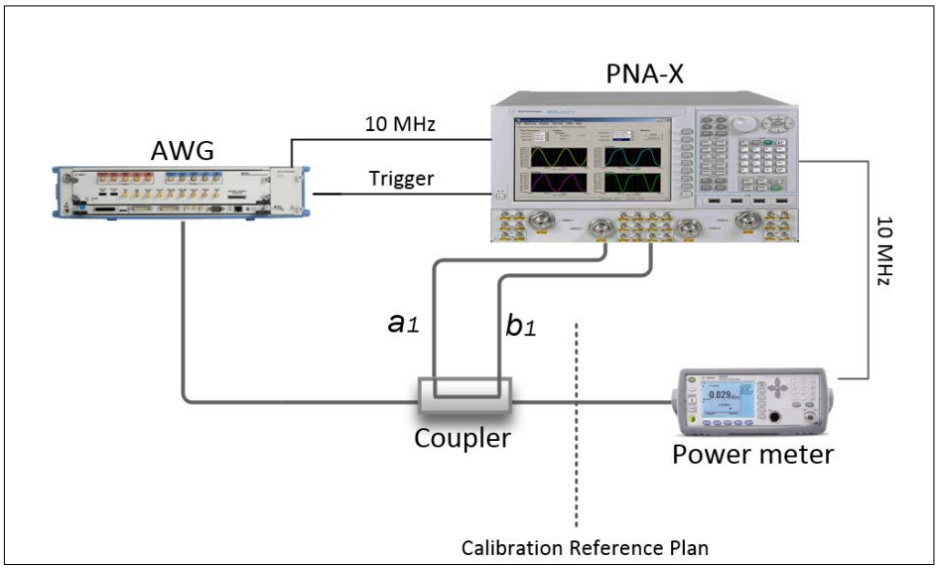


Figure 4.10 – Absolute Power Calibration

4.2.2.3 Phase Calibration

For linear measurement like S-parameter, the output frequency of the device is the same as the input frequency. As no new frequency components are created, the receivers do not have to measure outside of the excitation frequency. However, when the device studied is nonlinear, the measurement is more complex since the output spectrum contains spectral components at baseband, fundamental frequency and harmonics. To capture the full spectrum, the receiver Local Oscillator (LO) needs to move between these different frequencies. Unfortunately, the LO phase can randomly change between frequency sweeps and measurements which results in an inconsistent phase information. As a remedy to this problem, the PNA-X is used along with a comb generator which acts like a known nonlinearity that presents a constant phase reference to all signal harmonics. Indeed, all the receiver readings are rationed against the comb output to remove the phase error. Therefore, a nonlinear measurement for the four traveling waves need a fifth receiver connected to the comb generator.

In the case of multitones, intermodulation products are created on top of the harmonics. Therefore, the comb generator input frequency is not the fundamental frequency anymore, but should be a fraction of the frequency spacing between the multitones in order to ensure that the frequency grid at the comb generator output overlap with the intermodulation products and harmonics created. The major limitation to this technique is that the comb generator always creates harmonics of its input frequency with a decreasing spectral power. The higher the harmonic order is, the lower power it has. Therefore, lowering the comb generator input frequency as multitone measurement requires, means that the harmonics that overlap with the fundamental tones and intermodulation products will have low power and can be even below the noise floor. Consequently, this fact can degrade significantly the measurement quality and put a limit on the lowest possible spacing between the tones.

Taking into account the difficulties associated with multitones phase correction, relying in the frequency-swept mode to measure a modulated signals seems to be a cumbersome task. This is one of the reasons behind adopting a time domain capture of the travelling waves with the PNA-X. As signals are captured at once in the time domain then demodulated, the spectrum is generated by applying a Fourier transform where all the envelope spectral components are inherently phase coherent. The LO is kept fixed when measuring the envelope around the fundamental frequency, but must be changed each time a harmonic envelope is measured. As a result, the envelope spectral components of each harmonic are phase coherent relative to each other but the different envelopes

will be desynchronized. Therefore, a phase calibration is only needed when dealing with a harmonic behavioral model.

The DIOD Volterra model, up to this stage, is formulated just around the fundamental where the envelope is measured in a fixed LO manner. Consequently, no phase calibration is need to ensure relative phase coherency.

Chapter 5

Conclusion and Future Work

In this thesis, the benefits of constructing a behavioural model based on modulated signals for nonlinear unmatched devices were discussed. The availability of such models should benefit the overall accuracy of nonlinear circuit CAD simulators especially in system level simulations. Moreover, the strong connection between the modeling and characterization aspects was underlined. Indeed, the limitations as well as the advances in each aspect reflect similarly in the other. Any modeling effort needs to take into account the measurement capabilities. Therefore, this work describes the attempt to build a model based on the framework of the Volterra series using real life modulated signals and presents the characterization platform to extract it.

In the modeling part, the steps to derive the reduced complexity envelope domain Volterra model for two-port networks were presented. The model captures the behavior of the nonlinear unmatched device around fundamental frequency and does not include harmonics up to this point. The simulation based results were presented as initial validation and assessment of the model performance. As far as the characterization aspect is concerned, the progress towards building a characterization setup and its challenges were described. The under-development new characterization platform combines the measurement capabilities of an Agilent PNA-X and a hybrid load pull system to capture the vector-calibrated spectrum of the waveforms at the device input/output ports.

Future modeling work should include a comparison between the DIDO Volterra model performance and other measurement-based models like X-parameters and Cardiff Model. The comparison should focus in the interpolation and extrapolation capability of each model as well as the implementation complexity and measurement time. Furthermore, to fully capture the behavior of the DUT, the baseband and harmonics behavior should be captured on top of the fundamental frequency. Therefore, the actual model should be augmented with harmonic behaviour in a controlled fashion so that it does not reach impractical complexity. Finally, an important model improvement is its effective implementation in CAD simulator to enable its utilization in the design process.

On the other hand, future work in the characterization setup should be around finalizing the measurement platform. The remaining milestones include developing automation scripts, integrating the different building block of the measurement test-bench and writing validation routines. The characterization will enable the extraction the DIDO Volterra model from real measurements on transistors and RF power amplifier in unmatched environment but also any other envelope domain model. Besides, several enhancements can be applied to the platform following two major directions. Namely, high power handling capabilities since characterizing devices above 100 Watt is increasingly important and harmonic capabilities indispensable feature for transistors full characterization.

Bibliography

[1] P. Roblin, Nonlinear RF Circuits and Nonlinear Vector Network Analyzers, Cambridge University Press, 2002.

[2] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microwave Magazine,* vol. 10, pp. 52-64, 2009.

[3] S. S. Meena, C. Baylis, L. Dunleavy and M. Marbell, "Duty cycle dependent pulsed-IV simulation and thermal constant model fitting for LDMOS transistors," *ARFTG Microwave Measurement Symposium,* pp. 1-4, 2009.

[4] Microwave&RF, [Online]. Available: http://mwrf.com/test-and-measurement/three-probe-tuner-tackles-multiple-tasks. [Accessed 1 September 2014].

[5] F. M. Ghannouchi and S. M. Hashmi, "Load-Pull techniques and their Applications in Power Amplifiers Design," *IEEE BCTM,* 2011.

[6] S. Bonino, V. Teppati and A. Ferrero, "A novel methodology for fast harmonic-load control with a passive tuner and an active loop," *IEEE MTT-S Int. Microwave Symp. Dig.,* pp. 1158 - 1161, 2010.

[7] Y. Takayama, "A new load-pull characterization method for microwave power transistors," *IEEE MTT-S Int. Microwave Symp. Dig.,* pp. 218-220, 1967.

[8] G. P. Bava, U. Pisani and V. Pozzolo, "Active Load Technique for Load-Pull Characterisation at Microwave Frequencies," *Electronic Letters 18th* , vol. 18, no. 4, pp. 178-180, 1982.

[9] S. J. Hashim, M. S. Hashmi, T. Williams, S. Woodington, J. Benedikt and P. J. Tasker,, "Active Envelope Load-Pull for Wideband Multi-tone Stimulus Incorporating Delay Compensation," *Proc. 38th European Microwave Conf. (EuMC\'08),* pp. 317-320, 2008 .

[10] M. Marchetti, M. J. Pelk, K. Buisman, W. Neo, M. Spirito and . L. de Vreede, "Active Harmonic Load–Pull With Realistic Wideband Communications Signals," *IEEE Trans. Microwave Theory Techniques,* vol. 56, no. 12, pp. 2979 -2988, 2008.

[11] S. Doo, P. Roblin, V. Balasubramanian, R. Taylor, K. Dandu, J. Strahler, G. H. Jessen and J. Teyssier, "Pulsed active load-pull measurements for the design of high-efficiency class-B RF power amplifiers with GaN HEMTs," *IEEE Trans. Microwave Theory Techniques,* vol. 57, no. 4, pp. 881-889, 2009.

[12] Maury Microwave, "Hybrid-Active Load Pull," [Online]. Available: http://dialogs.maurymw.com/Extranet/96203/forms.aspx?msgid=f58d10ed-fea1-4985-84d4-4799d9baa800&LinkID=CH00096203eR00000029AD. [Accessed 10 August 2014].

[13] J. Verspecht and D. E. Root, "Poly-Harmonic Distorion Modeling," *IEEE Microwave Magazine,* 2006.

[14] Keysight Technologies, "Agilent 2-port and 4-port PNA-X Network Analyzer," 2014.

[15] G. Simpson, J. Hom, D. Gunyan and D. E. Root, "Load-Pull +NVNA=Enhanced X-Parameters for PA Designs with High Mismatch and Technology-Independent Large-Signal Device Models," *ARFTG Microwave Measuerment Symposium,* pp. 88-91, 2008.

[16] D. T. Bespalko and S. Boumaiza, "X-Parameters Measurement Challenges for Unmatched Device Characterization," *ARFTG Microwave Measuerment Symposium,* 2010.

[17] J. Verspecht, J. Horn, L. Betts, D. Gunyan, R. Pollard, C. Gillease and D. E. Root, "Extension of X-parameters to Include Long-Term Dynamic Memory Effects," *IEEE MTT-S Int. Microwave Symp. Dig.,* pp. 741-744, 2009.

[18] H. Qi, J. Benedikt and P. J. Tasker, "Data Utilization: From Direct Data Lookup to Behavioral Modeling," *IEEE Trans. Microwave Theory Techniques,* vol. 57, no. 6, pp. 1425-1432, 2009.

[19] S. Woodington, R. S. Saini, D. Williams, J. Lees, J. Benedikt and P. J. Tasker, "Behavioral model analysis of active harmonic load-pull measurements," *IEEE MTT-S Int. Microwave Symp. Dig.,* pp. 1688-1691, 2010.

[20] P. J. Tasker, "Practical Waveform Engineering," *IEEE Microwave Magazine,* vol. 10, no. 7, pp. 65-67, 2009.

[21] D. J. Williams and P. J. Tasker, "An automated active source and load pull measurement system," *6th IEEE High Frequency Postgraduate Student Colloquium,* 2001.

[22] M. Schetzen, The Volterra and Wiener Theories of Nonlinear Systems, Wiley, 1980.

[23] J. Kim and K. Konstantinou, "Digital predistortion of the wide-band signals based on power amplifier model with memory," *Electron. Lett.,* vol. 37, no. 23, pp. 1417-1418, 2001.

[24] A. Zhu, J. C. Pedro and T. J. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microwave Theory Techniques,* vol. 54, no. 12, pp. 4323-4332, 2006.

[25] A.-R. Amini, Analysis of Power Transistor Behavioural Modeling Techniques Suitable for Narrow-band Power amplifier Design, M.S. thesis, Dept. Elect. and Computer Eng., University of Waterloo, Waterloo, ON, 2012.

[26] B. Fehri and S. Boumaiza, "Baseband Equivalent Volterra Series for Behavioral Modeling and Digital Predistortion of Power Amplifiers Driven with Wideband Carrier Aggregated Signals," *IEEE Trans. Microwave Theory Techniques, accepted.*

[27] L. W. Naget, "SPICE2: A computer program to simulate semiconductor circuits," Berkeley, CA, 1975.

[28] K. S. Kundert and A. Sangiovanni-Vincentelh, "Simulation of nonlinear circuits in the frequency domain," *IEEE Trans. CAD,* Vols. CAD-5, pp. 521-535, 1986.

[29] E. Ngoya and R. Larcheveque, "Envelope transcient analysis: A new method for the transcient ans steady-state analysis of microwave communications circuits ans systems," *IEEE MTT-S Int. Microwave Symp. Dig.*, pp. 17-21, 1996.

[30] Keysight Technologies, "Circuit Envelope Simulation," 2006.

[31] Keysight Technologies, "M8190A 12 GSa/s Arbitrary Waveform Generator," [Online]. Available: http://www.keysight.com/en/pd-1969138-pn-M8190A/12-gsa-s-arbitrary-waveform-generator?cc=US&lc=eng. [Accessed 1 August 2014].

[32] J. P. Dunsomore, Handbook of Microwave Component Measurements: with Advanced VNA Techniques, Wiley, 2012.

[33] Keysight Technologies, "Applying Error Correction to Network Analyzer Measurements," 2002.

[34] Federal Institute of Metrology METAS, "METAS VNA Tools II," [Online]. Available: http://www.metas.ch/metasweb/Fachbereiche/Elektrizitaet/HF/VNATools/VNATools.html. [Accessed 1 August 2014].